A COMPREHENSIVE MODERN TREATMENT
OF ANALYTIC FUNCTION THEORY
AND ITS RAMIFICATIONS

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

VOLUME I

A. I. MARKUSHEVICH

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

VOLUME I

A. I. MARKUSHEVICH

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

Volume I

SELECTED RUSSIAN PUBLICATIONS IN THE MATHEMATICAL SCIENCES

Richard A. Silverman, *Editor*

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY., LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA (PRIVATE) LTD., *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

Volume I

A. I. MARKUSHEVICH

Professor of Mathematics
Moscow State University

Revised English Edition
Translated and Edited by

Richard A. Silverman

PRENTICE-HALL, INC.

Englewood Cliffs, N.J.

© 1965 by
PRENTICE-HALL, INC.
Englewood Cliffs, N.J.

All rights reserved. No part of this book may be reproduced in any form, by mimeograph or any other means, without permission in writing from the publisher.

Library of Congress Catalog Card Number 65–12876

Printed in the United States of America

91381 C

AUTHOR'S PREFACE

The theory of analytic functions of a complex variable is one of the outstanding accomplishments of classical mathematics. The rudiments of the theory had already begun to emerge in the eighteenth century, in the investigations of Euler and d'Alembert, but the full flowering of the subject was only possible in the nineteenth century, as a result of the fundamental researches of Cauchy, Riemann and Weierstrass. Its methods are widely used in many branches of mathematics, ranging from number theory to fluid dynamics, elasticity theory and electrodynamics. Even in our own day, when young scientists are so often preoccupied with more abstract branches of mathematics, the theory of analytic functions remains as significant as ever, both as a treasury of flexible and powerful analytic tools, and as a prototype or model of other theories dealing with generalizations of analytic functions in various directions.

It is the author's pleasant duty to cordially thank Dr. Richard A. Silverman, who translated the book into English, introduced many appropriate changes making the book easier to read, and enhanced the English-language edition with an abundance of problems and exercises.

This book stems from lectures given over the course of many years at Moscow University. In commending it to the attention of its new readership, the author would consider his purpose fulfilled if the present edition served to stimulate the interest of young mathematicians in one of the most remarkable branches of exact science, created by two centuries of joint effort on the part of

outstanding scholars of many nations, and if, at the same time, it represented a contribution by him and the translator towards the strengthening of cultural and scientific cooperation between their two countries.

A. I. M.

TRANSLATOR'S PREFACE

It would be difficult to recall in detail the ways in which the present version departs from the Russian original. Suffice it to say that in keeping with my philosophy of "enlightened translation," I thought it entirely appropriate to interact strongly with the initial text, especially since it lies within the area of my mathematical competence. This has led to a redistribution of the subject matter over three English volumes (the Russian original contains some 700 oversized pages, many in very small print), an overall restyling both of the typography and the art work (in particular, more theorems, examples, remarks, etc. have been explicitly identified as such), an adaptation of the bibliographic materials to suit the needs of those who do not know Russian (a class which presumably includes almost all the readers of this series), and so on. Certainly the most significant change is the addition of nearly 400 relevant problems, closely paralleling the text and stemming from a variety of sources, the most important of which is the excellent collection by Volkovyski, Lunts and Aramanovich (Moscow, 1961). The entire manuscript has also been read by Dr. T. J. Rivlin, who made many helpful suggestions.

Sections marked with asterisks can be omitted on first reading. However, it is only fair to say that they have been so marked merely because of their relative difficulty, and actually contain some of the most interesting material in the book.

R. A. S.

CONTENTS

PART 1 BASIC CONCEPTS, Page 1.

1 INTRODUCTION, Page 3.

1. Analytic Functions of a Real Variable, 3.
2. Infinitely Differentiable Functions, 5.
3. Motivation for Introducing the Complex Numbers. A Preview of Analytic Functions of a Complex Variable, 7.
Problems, 10.

2 COMPLEX NUMBERS, Page 12.

4. Geometric Representation of Complex Numbers, 12.
5. Complex Algebra, 14.
6. Powers and Roots of Complex Numbers, 17.
Problems, 19.

3 SETS AND FUNCTIONS. LIMITS AND CONTINUITY, Page 23.

7. Some Basic Definitions, 23.
8. Sequences of Complex Numbers. Limit Points and Limits of Sequences, 27.
9. Convergence of the Real and Imaginary Parts, Moduli and Arguments of a Complex Sequence, 32.
10. Series with Complex Terms, 35.
11. Limit Points of Sets. Bounded Sets, 42.

3 SETS AND FUNCTIONS, LIMITS AND CONTINUITY—Continued

12. The Limit of a Function of a Complex Variable, 44.
13. Continuous Functions. More Set Theory, 46.
14. The Distance between Two Sets, 52.
Problems, 54.

4 CONNECTEDNESS. CURVES AND DOMAINS, Page 59.

15. Connected Sets. Continuous Curves and Continua, 59.
16. Domains. Interior, Exterior and Boundary Points, 64.
17. Simply and Multiply Connected Domains, 66.
18. The Jordan Curve Theorem, 69.
19. Some Further Results, 72.
Problems, 74.

5 INFINITY AND STEREOGRAPHIC PROJECTION, Page 77.

20. Proper and Improper Complex Numbers, 77.
21. Stereographic Projection. Sets of Points on the Riemann Sphere, 79.
22. The Extended Complex Plane. The Point at Infinity, 83.
23. Conformality of Stereographic Projection. Continuous Curves in the Extended Plane, 87.
24. The Transformation $\zeta = 1/z$, 90.
25. Another Definition of an Angle with its Vertex at Infinity, 91
Problems, 92.

6 HOMEOMORPHISMS, Page 94.

26. The One-to-One Continuous Image of a Domain, 94.
27. Some Further Results, 99.
Problems, 100.

PART 2 DIFFERENTIATION. ELEMENTARY FUNCTIONS, Page 103.

7 DIFFERENTIATION AND THE CAUCHY-RIEMANN EQUATIONS, Page 105.

28. Derivatives and Differentials, 105.
29. Rules for Differentiating Functions of a Complex Variable, 107.
30. The Cauchy-Riemann Equations. Analytic Functions, 110.
Problems, 116.

8 GEOMETRIC INTERPRETATION OF THE DERIVATIVE. CONFORMAL MAPPING, Page 118.

31. Geometric Interpretation of Arg $f'(z)$, 118.
32. Geometric Interpretation of $|f'(z)|$, 121.
33. The Mapping $w = \dfrac{az + b}{cz + d}$, 122.
34. Conformal Mapping of the Extended Plane, 124.
Problems, 126.

9 ELEMENTARY ENTIRE FUNCTIONS, Page 128.

35. Polynomials, 128.
36. The Mapping $w = P_n(z)$, 130.
37. The Mapping $w = (z - a)^n$, 132.
38. The Exponential, 135.
39. The Mapping $w = e^z$, 140.
40. Some Functions Related to the Exponential, 146.
41. The Mapping $w = \cos z$, 150.
42. The Image of a Half-Strip under $w = \cos z$, 154.
Problems, 158.

10 ELEMENTARY MEROMORPHIC FUNCTIONS, Page 160.

43. Rational Functions, 160.
44. The Group Property of Möbius Transformations, 165.

10 ELEMENTARY MEROMORPHIC FUNCTIONS —Continued

45. The Circle-Preserving Property of Möbius Transformations, 168.
46. Fixed Points of a Möbius Transformation. Invariance of the Cross Ratio, 171.
47. Mapping of a Circle onto a Circle, 176.
48. Symmetry Transformations, 178.
49. Examples, 181.
* 50. Lobachevskian Geometry, 183.
51. The Mapping $w = \frac{1}{2}(z + \frac{1}{z})$, 197.
52. Transcendental Meromorphic Functions. Trigonometric Functions, 202.
Probems, 207.

11 ELEMENTARY MULTIPLE-VALUED FUNCTIONS, Page 212.

53. Single-Valued Branches. Univalent Functions, 212.
54. The Mapping $w = \sqrt[n]{z}$, 214.
55. The Mapping $w = \sqrt[n]{P(z)}$, 219.
56. The Logarithm, 224.
57. The Function z^a. Exponentials and Logarithms to an Arbitrary Base, 228.
58. The Mapping $w = \text{Arc cos } z$, 234.
59. The Mapping $w = z + \ln z$, 237.
Problems, 239.

PART 3 INTEGRATION. POWER SERIES, Page 243.

12 RECTIFIABLE CURVES. COMPLEX INTEGRALS, Page 245.

60. Some Basic Definitions, 245.
61. Integrals of Complex Functions, 248.
62. Properties of Complex Integrals, 250.
Problems, 253.

13 CAUCHY'S INTEGRAL THEOREM, Page 258.

63. A Preliminary Result, 258.
64. The Key Lemma, 259.
65. Proof of Cauchy's Integral Theorem, 261.
66. Application to the Evaluation of Definite Integrals, 272.
67. Cauchy's Integral Theorem for a System of Contours, 279.
68. Path-Independent Integrals. Primitives, 281.
69. The Integral as a Function of Its Upper Limit in a Multiply Connected Domain, 284.
Problems, 289.

14 CAUCHY'S INTEGRAL AND RELATED TOPICS, Page 293.

70. Cauchy's Integral Formula, 293.
71. Some Consequences of Theorem 14.1, 295.
72. Integrals of the Cauchy Type. Cauchy's Inequalities, 299.
* 73. Boundary Values of Integrals of the Cauchy Type, 306.
* 74. The Plemelj Formulas, 309.
Problems, 316.

15 UNIFORM CONVERGENCE. INFINITE PRODUCTS, Page 321.

75. Uniformly Convergent Series, 321.
76. Uniformly Convergent Sequences. Improper Integrals of the Cauchy Type, 330.
77. Infinite Products, 334.
Problems, 340.

16 POWER SERIES: RUDIMENTS, Page 344.

78. The Cauchy-Hadamard Formula, 344.
79. Taylor's Series. The Uniqueness Property, 348.

16 POWER SERIES: RUDIMENTS—Continued

80. The Relation between Power Series and Fourier Series, 353.
81. Expansion of an Analytic Function in Power Series, 358.
Problems, 365.

17 POWER SERIES: RAMIFICATIONS, Page 369.

82. The Interior Uniqueness Theorem. A-Points of Analytic Functions, 369.
83. The Maximum Modulus Principle and Some of Its Consequences. Lemniscates, 376.
84. Circular Elements. Regular and Singular Points, 382.
* 85. Behavior of a Power Series on Its Circle of Convergence, 395.
* 86. Compact Families of Analytic Functions, 411.
* 87. Vitali's Theorem. Analytic Functions Defined by Integrals, 417.
Problems, 423.

18 METHODS FOR EXPANDING FUNCTIONS IN TAYLOR SERIES, Page 429.

88. The Taylor Series of the Sum of a Series of Analytic Functions, 429.
89. The Taylor Series of a Composite Function, 432.
90. Division of Power Series, 436.
Problems, 444.

BIBLIOGRAPHY, Page 447.

INDEX, Page 451.

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

Volume I

Part 1

BASIC CONCEPTS

1

INTRODUCTION

1. Analytic Functions of a Real Variable

The mathematical discipline presented in this book has two names, the *theory of functions of a complex variable* and the *theory of analytic functions.* Each of these designations is justified by an aspect of the subject.

A function $f(x)$ of a real variable x, defined in some interval (a, b),[1] is said to be *analytic at the point* x_0 in (a, b) if in some neighborhood of x_0 we can represent $f(x)$ as the sum of a convergent power series in $x - x_0$:

$$f(x) = a_0 + a_1(x - x_0) + \cdots + a_n(x - x_0)^n + \cdots.$$

A function which is analytic at every point of (a, b) is said to be *analytic in the interval* (a, b). Any of the elementary functions studied in analysis is analytic on its whole domain of definition, except possibly at certain special points called *singular points.*

Example 1. The polynomial

$$P_m(x) = a_0 + a_1x + \cdots + a_mx^m \qquad (a_m \neq 0)$$

of degree m, the exponential function e^x and the trigonometric functions

[1] By (a, b) we mean the *open* interval $a < x < b$, and by $[a, b]$ the *closed* interval $a \leqslant x \leqslant b$. By a *neighborhood* of the point x_0 we mean an open interval of the form $x_0 - \varepsilon < x < x_0 + \varepsilon$ (see Sec. 8). In definitions, the locution "is said to be ... if" will be used with the meaning "is ... if and only if" (the dots indicate a missing predicate). A property is said to hold "on" a set of points E if it holds "at all points of" E, but when talking about intervals, neighborhoods, planes or half-planes, the word "in" is preferred to "on."

cos x, sin x are all analytic everywhere, since for any x_0 we have the expansions

$$P_m(x) = P_m(x_0) + \frac{P_m'(x_0)}{1!}(x - x_0) + \cdots + \frac{P_m^{(m)}(x_0)}{m!}(x - x_0)^m,$$

$$e^x = e^{x_0} + \frac{e^{x_0}}{1!}(x - x_0) + \cdots + \frac{e^{x_0}}{n!}(x - x_0)^n + \cdots,$$

$$\cos x = \cos x_0 - \frac{\sin x_0}{1!}(x - x_0) + \cdots + \frac{\cos(x_0 + \frac{1}{2}n\pi)}{n!}(x - x_0)^n + \cdots,$$

$$\sin x = \sin x_0 + \frac{\cos x_0}{1!}(x - x_0) + \cdots + \frac{\sin(x_0 + \frac{1}{2}n\pi)}{n!}(x - x_0)^n + \cdots.$$

Example 2. The function $1/(a - x)$ is analytic for $x \neq a$, since

$$\frac{1}{a - x} = \frac{1}{a - x_0}\frac{1}{1 - \dfrac{x - x_0}{a - x_0}}$$

$$= \frac{1}{a - x_0} + \frac{1}{(a - x_0)^2}(x - x_0) + \cdots + \frac{1}{(a - x_0)^{n+1}}(x - x_0)^n$$

if $x_0 \neq a$ and

$$\left|\frac{x - x_0}{a - x_0}\right| < 1.$$

Example 3. The function $x^{1/3}$ is analytic for $x \neq 0$, since

$$x^{1/3} = x_0^{1/3}\left(1 + \frac{x - x_0}{x_0}\right)^{1/3}$$

$$= x_0^{1/3} + \frac{1}{3}x_0^{-2/3}(x - x_0) + \cdots$$

$$+ \frac{1}{3}\left(\frac{1}{3} - 1\right)\cdots\left(\frac{1}{3} - n + 1\right)\frac{1}{n!}x_0^{\frac{1}{3}-n}(x - x_0)^n + \cdots$$

if $x_0 \neq 0$ and

$$\left|\frac{x - x_0}{x_0}\right| < 1.$$

Example 4. The function $\ln x$ is analytic on its whole domain of definition, since

$$\ln x = \ln x_0 + \ln\left(1 + \frac{x - x_0}{x_0}\right)$$

$$= \ln x_0 + \frac{1}{x_0}(x - x_0) - \frac{1}{2x_0^2}(x - x_0)^2 + \cdots$$

$$+ (-1)^{n-1}\frac{1}{nx_0^n}(x - x_0)^n + \cdots$$

if $x_0 > 0$ and

$$\left|\frac{x - x_0}{x_0}\right| < 1.$$

The fact that applying any of the basic algebraic and analytic operations (addition, subtraction, multiplication, division, differentiation and integration) to a convergent power series leads in general to another convergent power series[2] gives us some idea of the generality and importance of the class of analytic functions. Actually, the importance of analytic functions is even greater than these considerations suggest, since it turns out that all of the following functions are analytic (in appropriate intervals):

1. An analytic function of an analytic function;
2. An inverse function of an analytic function;
3. A function $f(x)$ satisfying an algebraic equation of degree n with analytic coefficients, i.e., an equation of the form

$$q_0(x) + q_1(x)f(x) + \cdots + q_n(x)[f(x)]^n = 0,$$

where $q_k(x)$, $k = 0, 1, \ldots, n$, are analytic functions;

4. A function $f(x)$ satisfying a linear differential equation of order n with analytic coefficients, i.e., an equation of the form

$$q_0(x) + q_1(x)\frac{df(x)}{dx} + \cdots + q_n(x)\frac{d^n f(x)}{dx^n} = g(x),$$

where $q_k(x)$, $k = 0, 1, \ldots, n$, and $g(x)$ are analytic functions.

Thus, it is hardly surprising that the most important classes of functions encountered in classical analysis and in its applications to problems of mechanics and physics consist of functions which are analytic everywhere except at certain *singular points*. This is the reason why it is so important to make a special study of the general properties of analytic functions.

2. Infinitely Differentiable Functions

Despite its generality, the class of analytic functions is only a proper subset (see p. 23) of the class of *infinitely differentiable functions* (i.e., functions with derivatives of all orders). The relation between the two classes is shown by the following

THEOREM 1.1. *A function $f(x)$ defined in a neighborhood of the point x_0 is analytic at x_0 if and only if*

1. *$f(x)$ is infinitely differentiable in a neighborhood of x_0;*

2. *There exist positive numbers δ and M such that for any x in the interval $(x_0 - \delta, x_0 + \delta)$ the inequality*

$$|f^{(k)}(x)| \leqslant M\frac{k!}{\delta^k} \tag{1.1}$$

holds for any nonnegative integer k.

---

[2] In the case of division, an exception occurs at the points where the divisor vanishes.

Proof. First we show that the condition (1.1) is necessary. If the function $f(x)$ is analytic at the point x_0, it can be represented as a power series

$$f(x) = a_0 + a_1(x - x_0) + \cdots + a_n(x - x_0)^n + \cdots \tag{1.2}$$

in some interval $(x_0 - \varepsilon, x_0 + \varepsilon)$. As we know from calculus,[3] the series (1.2) can be differentiated term by term any number of times, and in particular, the derivative of order k has the power series expansion

$$\begin{aligned} f^{(k)}(x) = k!\, a_k + \frac{(k+1)!\, a_{k+1}}{1!}(x - x_0) \\ + \frac{(k+2)!\, a_{k+2}}{2!}(x - x_0)^2 + \cdots, \end{aligned} \tag{1.3}$$

where $|x - x_0| < \varepsilon$. Let δ be such that $0 < 2\delta < \varepsilon$. Then the series (1.2) converges for $x - x_0 = 2\delta$, and hence

$$\lim_{n \to \infty} a_n(2\delta)^n = 0,$$

which implies that the sequence $\{a_n(2\delta)^n\}$ is bounded, i.e., that

$$|a_n(2\delta)|^n \leqslant M' \qquad (n = 0, 1, 2, \ldots). \tag{1.4}$$

We now use (1.3) and the inequality (1.4) to estimate $|f^{(k)}(x)|$ in the interval $(x_0 - \delta, x_0 + \delta)$, obtaining

$$\begin{aligned} |f^{(k)}(x)| &< k!\,|a_k| + \frac{(k+1)!\,|a_{k+1}|}{1!}\delta + \frac{(k+2)!\,|a_{k+2}|}{2!}\delta^2 + \cdots \\ &\leqslant k!\,\frac{M'}{(2\delta)^k} + \frac{(k+1)!}{1!}\frac{M'}{(2\delta)^{k+1}}\delta + \frac{(k+2)!}{2!}\frac{M'}{(2\delta)^{k+2}}\delta^2 + \cdots \\ &= \frac{k!M'}{2^k\delta^k}\left[1 + \frac{k+1}{1!}\frac{1}{2} + \frac{(k+1)(k+2)}{2!}\frac{1}{2^2} + \cdots\right] \\ &= \frac{k!M'}{2^k\delta^k}\left(1 - \frac{1}{2}\right)^{-(k+1)} = 2M'\frac{k!}{\delta^k}. \end{aligned} \tag{1.5}$$

To obtain the inequality (1.1), we need only set $2M' = M$ in (1.5).

Next we show that the condition (1.1) is sufficient. Let the function $f(x)$ be infinitely differentiable in the interval $(x_0 - \delta, x_0 + \delta)$, and suppose the inequality (1.1) is satisfied for x in $(x_0 - \delta, x_0 + \delta)$. Expanding $f(x)$ in a Taylor series with Lagrange remainder,[4] we obtain

$$\begin{aligned} f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \cdots \\ + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}, \end{aligned} \tag{1.6}$$

---

[3] See e.g., R. C. Buck, *Advanced Calculus*, McGraw-Hill Book Co., New York (1956), Theorem 24, p. 143.

[4] See e.g., D. V. Widder, *Advanced Calculus*, second edition, Prentice-Hall, Inc., Englewood Cliffs, N. J. (1961), p. 44.

where $0 < |\xi - x_0| < |x - x_0|$. Since

$$\left|\frac{f^{(n+1)}(\xi)}{(n+1)!}\right| |x - x_0|^{n+1} \leqslant \frac{M(n+1)!}{\delta^{n+1}(n+1)!} |x - x_0|^{n+1} = M\left(\frac{|x - x_0|}{\delta}\right)^{n+1},$$

the remainder term goes to zero as $n \to \infty$, provided that $|x - x_0| < \delta$. It follows that the function $f(x)$ can be written as a power series

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \cdots$$

in the interval $(x_0 - \delta, x_0 + \delta)$, i.e., $f(x)$ is an analytic function at the point x_0.

Remark. The analyticity criterion given by Theorem 1.1 leaves nothing to be desired from the standpoint of both rigor and explicitness. However, it is of little value either in applications or in theoretical investigations, since it is based on a knowledge of the behavior of derivatives of *all* orders in some neighborhood of the given point x_0 [see the inequality (1.1)].

3. Motivation for Introducing the Complex Numbers. A Preview of Analytic Functions of a Complex Variable

Since the square of any real number is nonnegative, even the simple quadratic equation

$$x^2 + 1 = 0 \tag{1.7}$$

has no real solutions (roots). However, it seems perfectly reasonable to require that any number system suitable for computational purposes allow us to solve the equation (1.7), or, for that matter, the general algebraic equation

$$a_0 + a_1 x + \cdots + a_n x^n = 0 \qquad (a_n \neq 0) \tag{1.8}$$

of degree n, where $a_0, a_1, \ldots, a_n$ are arbitrary real numbers. As already shown by (1.7), this goal can only be achieved if we somehow manage to extend the real number system, by making it part of another, larger number system. Since (1.7) is in a certain sense the "simplest" algebraic equation with no real roots, an obvious first approach to our problem is to introduce an "imaginary unit" $i = \sqrt{-1}$, and then consider "complex numbers" of the form

$$a + bi \quad \text{or} \quad a + ib, \tag{1.9}$$

where a and b are arbitrary real numbers, and algebraic operations are defined in the natural way, i.e., expressions of the form (1.9) are treated exactly like binomials $a + bx$ in an unknown x, except that the rule

$$i^2 = -1, \quad i^3 = -i, \quad i^4 = 1, \quad i^5 = i, \ldots \tag{1.10}$$

is used to eliminate all powers of i higher than the first. If this is done, the roots of (1.7) are the particularly simple complex numbers i and $-i$. Moreover, setting $b = 0$, we see that the numbers of the form (1.9) include the real numbers as a special case (an essential feature).

Surprisingly enough, as we shall see later (cf. Theorem 17.7 and Problem 17.14), it turns out that once we allow x to take complex values, equation (1.8) always has a root, even if the coefficients $a_0, a_1, \ldots, a_n$ are themselves complex numbers, a result known as the *fundamental theorem of algebra.* In this sense, the algebraic rewards for introducing complex numbers of the form (1.9) are all one might hope for. As we now indicate, the analytic rewards for taking this step are also very far-reaching.

Following the French mathematician Cauchy (1789–1857), suppose we try to develop a theory of analytic functions of a *complex variable*, i.e., of a variable of the form

$$z = x + iy$$

where x and y are two real variables. The required generalization of the concept of an analytic function of a real variable is simple and natural. In fact, we need only generalize the limit concept to a series with complex terms (see Sec. 10) and then observe that a series

$$a_0 + a_1(x - x_0) + \cdots + a_n(x - x_0)^n + \cdots \tag{1.11}$$

which converges for all values of the real variable x satisfying the condition $|x - x_0| < \varepsilon$ continues to converge for all values of the complex variable $z = x + iy$ satisfying the condition[5]

$$|z - x_0| = \sqrt{(x - x_0)^2 + y^2} < \varepsilon.$$

In this sense, the series

$$a_0 + a_1(z - x_0) + \cdots + a_n(z - x_0)^n + \cdots .$$

defines a *function of a complex variable* which can be regarded as the *extension* or *continuation* of the analytic function (1.11) of the real variable x into the "complex domain."

This idea of continuing a real function into the complex domain had been used long before the time of Cauchy. One example has already been noted in connection with the fundamental theorem of algebra, where polynomials are defined for complex values of the unknown (no problem of convergence arises in this case). Another example is afforded by the celebrated formula

$$e^{ix} = \cos x + i \sin x,$$

due to Euler (1707–1783), relating the exponential and trigonometric functions. Thus, Cauchy's historic contribution did not merely consist in

[5] Here, for the purposes of our qualitative discussion, we anticipate certain results which will be given a rigorous treatment in due course.

writing a complex variable z in place of a real variable x in power series, as had already been done in the eighteenth century. Rather, his contribution was to lay the foundations for a systematic theory of functions of a complex variable, constructed by analogy with the classical theory of functions of a real variable, and including a theory of limits, continuity, derivatives, integrals and series, all involving complex functions. In carrying out this program, the following fundamental fact came to light:

THEOREM 1.2.[6] *A function $f(z)$ of a complex variable is analytic on a domain G, i.e., $f(z)$ can be represented in the form*[7]

$$f(z) = a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots$$

in a neighborhood of every point z_0 in G, if and only if $f(z)$ is differentiable at every point z_0 of G. In other words, if $f(z)$ has a first derivative on G, it has a power series expansion at every point of G, and conversely.

As we saw in Sec. 2 (cf. Theorem 1.1), the connection between differentiability and analyticity of a function is quite complicated if we confine ourselves to real values of the independent variable. The reason why the situation becomes much simpler when we make the transition to functions of a complex variable is essentially that the existence of a derivative implies a great deal about a function when the independent variable has two "degrees of freedom" (see Chap. 7). The study of functions of a complex variable has many other advantages over the study of functions defined only for real values of their arguments. In particular, it is found that the analyticity of a complex function has a simple characterization in terms of the behavior of *integrals* of $f(z)$ (see Chap. 13). It also turns out that complex variable methods illuminate certain otherwise puzzling aspects of real variable theory. For example, the two real power series

$$\frac{1}{1 - x} = 1 + x + x^2 + \cdots,$$

and

$$\frac{1}{1 + x^2} = 1 - x^2 + x^4 - \cdots \tag{1.12}$$

both converge for $|x| < 1$ and diverge for $|x| \geq 1$. In the first case, this behavior seems natural since $(1 - x)^{-1}$ becomes infinite for $x = 1$, but in the second case, it seems strange since $(1 + x^2)^{-1}$ is finite for all real x. However, if we allow x to take *complex* values, $(1 + x^2)^{-1}$ becomes infinite for $x = \pm i$, and in fact this does explain why the second series diverges for $|x| > 1$.[8]

---

[6] Theorem 1.2 is a consequence of Theorems 16.4 and 16.7. See also Remark 1, p. 363.

[7] In general, $a_0, a_1, \ldots, a_n, \ldots$ z and z_0 are complex, but they may be real in special cases.

[8] See Example 1, p. 392.

For these and many other reasons that will become apparent later, the study of *analytic functions of a complex variable*, to which this book is devoted, is of fundamental importance in both pure and applied mathematics (cf. the opening sentence of Sec. 1).

PROBLEMS

1.1. Verify that the functions $\cosh x$ and $e^{\sin x}$ are analytic for all x.

1.2. Find the points at which the functions $\sec x$ and $\cot x$ fail to be analytic.

1.3. For each of the functions

$$e^x, \quad e^{x^2}, \quad e^{e^x},$$

determine a constant M such that

$$|f^{(k)}(x)| \leqslant Mk! \qquad (k = 0, 1, 2, \ldots)$$

for any x in the interval $(-1, 1)$.

1.4. Verify that the function

$$\begin{aligned} f(x) &= e^{-1/x^2} && \text{for} \quad |x| > 0, \\ f^{(k)}(0) &= 0 && \text{for} \quad k = 0, 1, 2, \ldots \end{aligned}$$

is infinitely differentiable but not analytic at the point $x = 0$. Discuss the analyticity of $f(x)$ at nonzero points.

1.5. Verify that the function $f(x)$ of the preceding problem fails to satisfy the inequality (1.1) in any neighborhood of the origin for any M.

1.6. Verify that the function

$$g(x) = \begin{cases} e^{-1/x^2} & \text{for} \quad x > 0, \\ 0 & \text{for} \quad x \leqslant 0 \end{cases}$$

is infinitely differentiable but not analytic at the point $x = 0$.

1.7. Construct an infinitely differentiable function which vanishes outside a given finite interval $[a, b]$.

Hint. Consider Problem 1.6.

1.8. Construct an infinitely differentiable function which is zero for $x \leqslant a$ and 1 for $x \geqslant b$ $(b > a)$.

Hint. Integrate the result of Problem 1.7.

1.9. Prove the following theorem due to Bernstein: If $f(x)$ is infinitely differentiable in a neighborhood of x_0, and if

$$f^{(k)}(x) \geqslant 0 \qquad (k = 0, 1, 2, \ldots)$$

for all x in this neighborhood, then $f(x)$ is analytic at x_0 (and in fact in the whole neighborhood).

Hint. If x belongs to the given neighborhood of x_0, say $(x_0 - \varepsilon, x_0 + \varepsilon)$, choose a and b such that $x_0 - \varepsilon < a < x < b < x_0 + \varepsilon$. Writing the remainder term of the Taylor series (1.6) in the form[9]

$$\begin{aligned} R_n(x) &= \frac{1}{n!}\int_a^x f^{(n+1)}(t)(x - t)^n \, dt \\ &= \frac{(x - a)^n}{n!}\int_0^1 f^{(n+1)}[(x - a)u + a](1 - u)^n \, du, \end{aligned}$$

show that

$$R_n(x) \leqslant \left(\frac{x - a}{b - a}\right)^{n+1} R_n(b) \leqslant \left(\frac{x - a}{b - a}\right)^{n+1} f(b).$$

---

[9] See e.g., D. V. Widder, *op. cit.*, p. 43.

2

COMPLEX NUMBERS

4. Geometric Representation of Complex Numbers

As already noted in Sec. 3, by a *complex number* we mean an expression of the form $a + ib$, where a and b are real numbers and i is the "imaginary unit." If $c = a + ib$, a is called the *real part* of c, written Re c, and b is called the *imaginary part* of c, written Im c. By the complex number *zero*, we mean the number $0 = 0 + i0$, with zero real and imaginary parts. By definition, two complex numbers c_1 and c_2 are *equal* if and only if

$$\text{Re } c_1 = \text{Re } c_2, \qquad \text{Im } c_1 = \text{Im } c_2.$$

If Im $c = 0$, $c = a + ib$ reduces to a real number, while if Im $c \neq 0$, c is said to be *imaginary* and if Re $c = 0$, Im $c \neq 0$, c is said to be *purely imaginary*.

Complex numbers can be represented geometrically as points in the plane, a fact which is not only useful but virtually indispensable. Introducing a rectangular coordinate system in the plane, we can identify the complex number $z = x + iy$ with the point $P = (x, y)$, as shown in Figure 2.1. In this way, we establish a one-to-one correspondence between the set of all complex numbers and the set of all points in the plane. Clearly, under this mapping, the set of all real numbers corresponds to the x-axis (hence called the *real axis*)

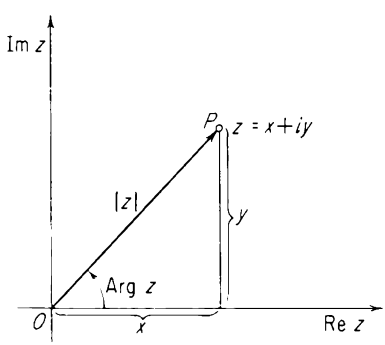

FIGURE 2.1

and the set of all purely imaginary numbers corresponds to the y-axis (hence called the *imaginary axis*), while the set of all imaginary numbers corresponds to all points which do not lie on the real axis. Moreover, the complex number 0 corresponds to the point of intersection of the x-axis and y-axis, i.e., the origin of coordinates. The plane whose points represent the complex numbers is called the *complex plane*, or the *z-plane*, *w-plane*, ..., depending on the letter z, w, ... used to denote a generic complex number. With the understanding that such a complex plane has been constructed, the terms "complex number $x + iy$" and "point $x + iy$" will be used interchangeably.

Another entirely equivalent way of representing the complex number $z = x + iy$ is to use the vector $\vec{OP}$ joining the origin O of the complex plane to the point $P = (x, y)$, instead of using the point P itself (see Figure 2.1). The length of the vector $\vec{OP}$ is called the *modulus* or *absolute value* of the complex number z, and is denoted by $|z|$. The angle between the positive real axis and the vector $\vec{OP}$, more exactly, the angle through which the positive real axis must be rotated to cause it to have the same direction as $\vec{OP}$ (considered positive if the rotation is counterclockwise and negative otherwise), is called the *argument* of the complex number z, and is denoted by Arg z.[1] In other words, $|z|$ and Arg z are the polar coordinates r and Φ of the point with rectangular coordinates x and y, i.e.,

$$x = \operatorname{Re} z = r \cos \Phi, \qquad y = \operatorname{Im} z = r \sin \Phi.$$

It follows at once that

$$z = x + iy = r(\cos \Phi + i \sin \Phi), \tag{2.1}$$

where (2.1) is called the *trigonometric form* of the complex number a.

Clearly, the quantity Arg z is defined only to within an integral multiple of 2π. However, there is one and only one value of Arg z, say φ, which satisfies the inequality

$$-\pi < \varphi \leqslant \pi,$$

and we shall call φ the *principal value* of the argument z, written arg z. The relation between Arg z and arg z is given by

$$\operatorname{Arg} z = \arg z + 2n\pi,$$

where n ranges over all the integers $0, \pm 1, \pm 2, \ldots$. It is an immediate consequence of Figure 2.1 and our definitions that[2]

$$|z| = \sqrt{x^2 + y^2}$$

and

$$\tan (\arg z) = \frac{y}{x}. \tag{2.2}$$

---

[1] In talking about the vector $\vec{OP}$, we have tacitly assumed that P is not the origin, i.e., that $z \neq 0$. If $z = 0$, then $|z| = 0$ and Arg z is indeterminate.

[2] By $\sqrt{a}$, where $a > 0$, we shall always mean the *positive* square root of a.

Some care is required in inverting (2.2), since the arc tangent of a real number x, written Arc tan x, is only defined to within an integral multiple of π. However, there is one and only one value of Arc tan x, say α, which satisfies the inequality

$$-\frac{\pi}{2} < \alpha \leqslant \frac{\pi}{2},$$

and we shall call α the *principal value* of the arc tangent of x, written arc tan x. We can now invert the relation (2.2), obtaining

$$\arg z = \begin{cases} \text{arc tan}\,\dfrac{y}{x} & \text{if } x > 0, \\ \text{arc tan}\,\dfrac{y}{x} + \pi & \text{if } x < 0,\ y \geqslant 0, \\ \text{arc tan}\,\dfrac{y}{x} - \pi & \text{if } x < 0,\ y < 0. \end{cases}$$

Moreover, when y/x becomes infinite, we clearly have

$$\arg z = \begin{cases} \dfrac{\pi}{2} & \text{if } x = 0,\ y > 0, \\ -\dfrac{\pi}{2} & \text{if } x = 0,\ y < 0, \end{cases}$$

while the case $z = 0$ (i.e., $x = y = 0$) is indeterminate, as already noted.

The complex numbers $x + iy$ and $x - iy$ are said to be *conjugate complex numbers* or *complex conjugates* (of each other). If one of these numbers is denoted by z, the other is denoted by $\bar{z}$. Obviously, the points z and $\bar{z}$ are symmetric with respect to the real axis, and

$$|\bar{z}| = |z|, \qquad \overline{(\bar{z})} = z.$$

Moreover,

$$\arg z = -\arg \bar{z},$$

unless z is a negative real number, in which case

$$\arg z = \arg \bar{z} = \pi.$$

5. Complex Algebra

It follows from the prescription of Sec. 3 [cf. (1.10)] that the sum and product of two complex numbers $z_1 = x_1 + iy_1$, $z_2 = x_2 + iy_2$ are given by

$$z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2), \tag{2.3}$$

$$z_1 z_2 = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1), \tag{2.4}$$

while the difference and quotient of z_1 and z_2 are given by

$$z_1 - z_2 = (x_1 - x_2) + i(y_1 - y_2),$$
$$\frac{z_1}{z_2} = \frac{x_1x_2 + y_1y_2}{x_2^2 + y_2^2} + i\frac{x_2y_1 - x_1y_2}{x_2^2 + y_2^2} \qquad (x_2^2 + y_2^2 \neq 0).$$

If z_1, z_2 and z_3 are any three complex numbers, the relations (2.3) and (2.4) imply that

$$z_1 + z_2 = z_2 + z_1,$$
$$z_1z_2 = z_2z_1$$

(addition and multiplication are *commutative*),

$$(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3),$$
$$(z_1z_2)z_3 = z_1(z_2z_3)$$

(addition and multiplication are *associative*), and

$$z_1(z_2 + z_3) = z_1z_2 + z_1z_3$$

(multiplication is *distributive* with respect to addition). Moreover, the product of two complex numbers is zero if and only if at least one of the factors is zero, any complex number z has a unique negative $-z$ such that $z + (-z) = 0$, and any nonzero complex number z has a unique inverse $z^{-1} = 1/z$ such that $zz^{-1} = 1$. (Verify these statements.) In other words, the complex numbers satisfy all the axioms of a *field*.[3]

Guided by (2.3) and (2.4), one might define the complex numbers (without recourse to the "imaginary unit" i) as ordered pairs

$$(x_1, y_1), (x_2, y_2), \ldots$$

of real numbers $x_1, y_1, x_2, y_2, \ldots$, where two pairs (x_1, y_1) and (x_2, y_2) are equal if and only if $x_1 = x_2$, $y_1 = y_2$, while addition and multiplication of pairs are defined by

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2),$$
$$(x_1, y_1)(x_2, y_2) = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1).$$

With this approach, the relation $i^2 = -1$ has the analogue

$$(0, 1)(0, 1) = (-1, 0),$$

which is sometimes adduced as an "explanation" of the "imaginary unit" i. However, it should be noted that the extension of the real numbers to the complex numbers is no more "mysterious" than the extension of the integers

---

[3] See e.g., G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, revised edition, The Macmillan Co., New York (1953), p. 35.

to the rational numbers, and less complicated than the extension of the rational numbers to the real numbers.[4]

Geometrically, addition of two complex numbers z_1 and z_2 corresponds to addition of the vectors representing them [see Figure 2.2(a)]. Similarly, the difference $z_1 - z_2$ is represented by the vector joining the point z_1 to the point z_2 [see Figure 2.2(b)]. It follows that $\rho(z_1, z_2)$, the distance between the two points z_1 and z_2 equals the modulus of the complex number $z_1 - z_2$:

$$\rho(z_1, z_2) = |z_1 - z_2|.$$

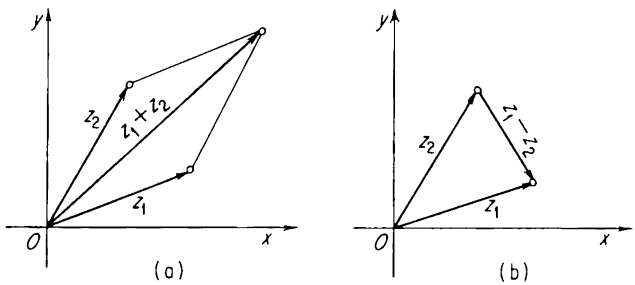

FIGURE 2.2

Then the familiar fact that the length of one side of a triangle cannot exceed the sum of the lengths of the other two sides implies the following two inequalities involving complex numbers:

$$|z_1 + z_2| \leqslant |z_1| + |z_2|, \tag{2.5}$$

$$|z_1 - z_2| \geqslant ||z_1| - |z_2||. \tag{2.6}$$

The relation (2.5) can be generalized immediately to the case of n complex numbers:

$$|z_1 + z_2 + \cdots + z_n| \leqslant |z_1| + |z_2| + \cdots + |z_n|. \tag{2.7}$$

In (2.5), (2.6) and (2.7), equality occurs if and only if all the (nonzero) complex numbers have the same argument. We note that according to (2.4), if $z = x + iy$, then

$$z\bar{z} = x^2 + y^2 = |z|^2,$$

and hence

$$|z| = \sqrt{x^2 + y^2}$$

and

$$|z_1 + \cdots + z_n| = \sqrt{(z_1 + \cdots + z_n)(\bar{z}_1 + \cdots + \bar{z}_n)}.$$

[4] See e.g., G. Birkhoff and S. MacLane, *op. cit.*, Chaps. 2 and 4.

To give a geometric interpretation of multiplication and division of two complex numbers z_1 and z_2, we first write z_1 and z_2 in trigonometric form [cf. (2.1)], i.e.,

$$z_1 = r_1(\cos \Phi_1 + i \sin \Phi_1),$$
$$z_2 = r_2(\cos \Phi_2 + i \sin \Phi_2),$$

where

$$r_1 = |z_1|, \qquad \Phi_1 = \operatorname{Arg} z_1,$$
$$r_2 = |z_2|, \qquad \Phi_2 = \operatorname{Arg} z_2.$$

Then, according to (2.4),

$$\begin{aligned} z_1 z_2 &= r_1 r_2[(\cos \Phi_1 \cos \Phi_2 - \sin \Phi_1 \sin \Phi_2) \\ &\quad + i(\sin \Phi_1 \cos \Phi_2 + \cos \Phi_1 \sin \Phi_2)] \\ &= r_1 r_2[\cos (\Phi_1 + \Phi_2) + i \sin (\Phi_1 + \Phi_2)], \end{aligned} \tag{2.8}$$

and therefore

$$|z_1 z_2| = |z_1|\,|z_2|,$$
$$\operatorname{Arg} z_1 z_2 = \operatorname{Arg} z_1 + \operatorname{Arg} z_2.$$

The last formula means that the set of values of $\operatorname{Arg} z_1 z_2$ is obtained by forming all possible sums of a value of $\operatorname{Arg} z_1$ and a value of $\operatorname{Arg} z_2$. Thus, the geometric interpretation of multiplication of z_1 by z_2 ($z_1 \neq 0$, $z_2 \neq 0$) is the following: Draw the vector z_1 (with initial point at the origin), multiply its length by the factor $|z_2|$,[5] and rotate the resulting vector counterclockwise through the angle $\operatorname{Arg} z_2$. Similarly, if $z_2 \neq 0$, it is easily verified that

$$\begin{aligned} \left|\frac{z_1}{z_2}\right| &= \frac{|z_1|}{|z_2|}, \\ \operatorname{Arg} \frac{z_1}{z_2} &= \operatorname{Arg} z_1 - \operatorname{Arg} z_2, \end{aligned} \tag{2.9}$$

with a corresponding geometric interpretation of division. It follows from the second of the formulas (2.9) that $\operatorname{Arg} (z_1/z_2)$ is the angle between the vectors z_2 and z_1, measured from z_2 to z_1 in the counterclockwise direction.

6. Powers and Roots of Complex Numbers

Let n be a positive integer and z a complex number. Then by the nth *power* of z, written z^n, we mean z multiplied by itself n times, just as in the case of real numbers. If $z = 0$, we obviously have $z^n = 0$. If $z \neq 0$, then, setting

$$z = r(\cos \Phi + i \sin \Phi) \tag{2.10}$$

and repeatedly applying (2.8), we find that

$$z^n = r^n(\cos n\Phi + i \sin n\Phi). \tag{2.11}$$

[5] This can correspond to either stretching or shrinking (or neither, if $|z_2| = 1$).

When $r = 1$, (2.10) and (2.11) together imply the following formula, known as *De Moivre's theorem*:

$$(\cos \Phi + i \sin \Phi)^n = \cos n\Phi + i \sin n\Phi. \tag{2.12}$$

Using the binomial theorem to expand the left-hand side of (2.12), and equating real and imaginary parts of the resulting equation, we can express $\cos n\Phi$ or $\sin n\Phi$ entirely in terms of powers of $\cos \Phi$ and $\sin \Phi$. For the same algebraic reasons as in the case of real numbers, we define

$$z^0 = 1, \qquad z^{-n} = \frac{1}{z^n}.$$

Then formulas (2.11) and (2.12) are valid for all integers $n = 0, \pm 1, \pm 2, \ldots$.

Next, we consider the problem of extracting the root of a complex number. Again, let n be a positive integer and z a complex number. Then by the nth *root* of z, written $\sqrt[n]{z}$ or $z^{1/n}$, we mean any complex number ζ satisfying the equation

$$\zeta^n = z.$$

If $z = 0$, we obviously have $\zeta = 0$. If $z \neq 0$, then, setting

$$\zeta = \rho(\cos \Theta + i \sin \Theta), \qquad z = r(\cos \Phi + i \sin \Phi),$$

we find that

$$\rho^n(\cos n\Theta + i \sin n\Theta) = r(\cos \Phi + i \sin \Phi).$$

It follows that[6]

$$\rho^n = r, \qquad \rho = \sqrt[n]{r} = \sqrt[n]{|z|}$$

and

$$n\Theta = \Phi + 2m\pi \qquad (m = 0, \pm 1, \pm 2, \ldots),$$

i.e.,

$$n\Theta = \operatorname{Arg} z, \qquad \Theta = \frac{\operatorname{Arg} z}{n}.$$

Therefore we have

$$\sqrt[n]{z} = \sqrt[n]{|z|}\left(\cos \frac{\operatorname{Arg} z}{n} + i \sin \frac{\operatorname{Arg} z}{n}\right). \tag{2.13}$$

It is clear from (2.13) that values of $\operatorname{Arg} z$ which differ by $2k\pi$, where k is an integer not divisible by n, correspond to different values of $\sqrt[n]{z}$. Accordingly, when $\operatorname{Arg} z$ runs through the values

$$\arg z, \arg z + 2\pi, \ldots, \arg z + 2(n-1)\pi, \tag{2.14}$$

we obtain n different values of $\sqrt[n]{z}$. Since any other value of $\operatorname{Arg} z$ differs from one of the numbers (2.14) by a multiple of $2n\pi$, these n values of $\sqrt[n]{z}$

[6] By $\sqrt[n]{a} = a^{1/n}$, where $a > 0$ and n is a positive integer, we shall always mean the unique *positive* nth root of a (cf. footnote 2, p. 13).

exhaust all the possibilities. Thus, the nth root of z has n different values (if $z \neq 0$), which are all implicit in the formula (2.13). The value of $\sqrt[n]{z}$ equal to

$$\sqrt[n]{|z|}\left(\cos\frac{\arg z}{n} + i\sin\frac{\arg z}{n}\right)$$

is called the *principal value* of $\sqrt[n]{z}$, since when z is a positive real number, the principal value of $\sqrt[n]{z}$ coincides with the *positive* nth root of z (cf. footnote 6, p. 18). It follows from the formula

$$\operatorname{Arg} z = \arg z + 2m\pi = \arg z + \operatorname{Arg} 1$$

that

$$\begin{aligned}\sqrt[n]{z} &= \sqrt[n]{|z|}\left(\cos\frac{\arg z + \operatorname{Arg} 1}{n} + i\sin\frac{\arg z + \operatorname{Arg} 1}{n}\right)\\ &= \sqrt[n]{|z|}\left(\cos\frac{\arg z}{n} + i\sin\frac{\arg z}{n}\right)\left(\cos\frac{\operatorname{Arg} 1}{n} + i\sin\frac{\operatorname{Arg} 1}{n}\right),\end{aligned}$$

i.e., all the values of $\sqrt[n]{z}$ can be obtained from the principal value by multiplying the latter by the different nth roots of unity.

Finally, we define an arbitrary rational power of a complex number by the formula

$$z^{m/n} = (\sqrt[n]{z})^m,$$

where m and $n > 0$ are relatively prime integers. Then we have

$$\begin{aligned}z^{m/n} &= \left[\sqrt[n]{|z|}\cos\left(\frac{\operatorname{Arg} z}{n} + i\sin\frac{\operatorname{Arg} z}{n}\right)\right]^m\\ &= (\sqrt[n]{|z|})^m\left(\cos\frac{m\operatorname{Arg} z}{n} + i\sin\frac{m\operatorname{Arg} z}{n}\right)\\ &= |z|^{m/n}\left(\cos\frac{m\operatorname{Arg} z}{n} + i\sin\frac{m\operatorname{Arg} z}{n}\right),\end{aligned}$$

where $|z|^{m/n}$ is defined as the positive number $(\sqrt[n]{|z|})^m = \sqrt[n]{|z|^m}$.

PROBLEMS

2.1. Represent the following complex numbers in trigonometric form:

a) $1 + i$; b) $-1 + i$; c) $-1 - i$; d) $1 - i$; e) $1 + i\sqrt{3}$; f) $-1 + i\sqrt{3}$; g) $-1 - i\sqrt{3}$; h) $1 - i\sqrt{3}$; i) $\sqrt{3} - i$; j) $2 + \sqrt{3} + i$.

2.2. Calculate the following expressions:

a) $\dfrac{1 + i\tan\alpha}{1 - i\tan\alpha}$; b) $\dfrac{(1 + 2i)^3 - (1 - i)^3}{(3 + 2i)^3 - (2 + i)^2}$; c) $\dfrac{(1 - i)^5 - 1}{(1 + i)^5 + 1}$;

d) $\dfrac{(1 + i)^9}{(1 - i)^7}$.

2.3. Calculate

a) $(a + b\omega + c\omega^2)(a + b\omega^2 + c\omega)$; b) $(a + b)(a + b\omega)(a + b\omega^2)$;
c) $(a + b\omega + c\omega^2)^3 + (a + b\omega^2 + c\omega)^3$; d) $(a\omega^2 + b\omega)(b\omega^2 + a\omega)$,

where $\omega = -\frac{1}{2} + \frac{1}{2}\sqrt{3}i$.

Ans. b) $a^3 + b^3$; d) $a^2 - ab + b^2$.

2.4. Find the complex numbers which are conjugate to a) their own squares, b) their own cubes.

Ans. b) $0, 1, i, -1, -i$.

2.5. Suppose that a complex number u is obtained as a result of a finite number of rational operations (i.e., addition, subtraction, multiplication and division) applied to the complex numbers $z_1, z_2, \ldots, z_n$. Prove that the same operations applied to the complex conjugates $\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_n$ leads to the number $\bar{u}$ which is the complex conjugate of u.

2.6. Defining rational numbers as ordered pairs of integers, write the corresponding rules for addition and multiplication of rational numbers. Verify that addition and multiplication are commutative and associative, and that multiplication is distributive with respect to addition.

Ans. $(x_1, y_1) + (x_2, y_2) = (x_1y_2 + x_2y_1, y_1y_2)$, $(x_1, y_1)(x_2, y_2) = (x_1x_2, y_1y_2)$, where $y_1 \neq 0$, $y_2 \neq 0$.

2.7. Prove the inequalities

$$|z_1 + z_2| \leqslant |z_1| + |z_2|,$$
$$|z_1 - z_2| \geqslant ||z_1| - |z_2||$$

algebraically.

2.8. Prove the identity

$$|z_1 + z_2|^2 + |z_1 - z_2|^2 = 2(|z_1|^2 + |z_2|^2),$$

and interpret it geometrically.

2.9. By a purely geometric argument, prove that

$$|z - 1| \leqslant ||z| - 1| + |z||\arg z|.$$

2.10. Solve the following equations:

a) $|z| - z = 1 + 2i$; b) $|z| + z = 2 + i$.

Ans. a) $z = \frac{3}{2} - 2i$.

2.11. Prove that every complex number (except -1) of unit modulus can be represented in the form

$$z = \frac{1 + it}{1 - it},$$

where t is a real number.

2.12. Prove that

$$|z_1| + |z_2| = \left|\frac{z_1 + z_2}{2} - u\right| + \left|\frac{z_1 + z_2}{2} + u\right|,$$

where z_1 and z_2 are any two complex numbers, and $u = \sqrt{z_1z_2}$.

2.13. Prove that

$$|(1 + i)z^3 + iz| < \tfrac{3}{4},$$

if $|z| < \frac{1}{2}$.

2.14. Prove that

$$\arg \frac{z_3 - z_2}{z_3 - z_1} = \frac{1}{2} \arg \frac{z_2}{z_1},$$

if $|z_1| = |z_2| = |z_3|$.

2.15. Find a necessary and sufficient condition for three distinct points z_1, z_2 and z_3 to lie on the same line.

2.16. Suppose the points $z_1, z_2, \ldots, z_n$ all lie on one side of a line drawn through the origin of the complex plane. Prove that the same is true of the points $1/z_1, 1/z_2, \ldots, 1/z_n$. Moreover, show that

$$z_1 + z_2 + \cdots + z_n \neq 0, \qquad \frac{1}{z_1} + \frac{1}{z_2} + \cdots + \frac{1}{z_n} \neq 0.$$

2.17. Prove that if $z_1 + z_2 + z_3 = 0$ and $|z_1| = |z_2| = |z_3| = 1$, then the points z_1, z_2 and z_3 are the vertices of an equilateral triangle inscribed in the unit circle $|z| = 1$.

2.18. What are the loci of the points z which satisfy the following relations:

a) $|z - z_1| = |z - z_2|$; b) $|z - 2| + |z + 2| = 5$;

c) $|z - 2| - |z + 2| > 3$; d) $0 < \operatorname{Re}(iz) < 1$;

e) $|z| = \operatorname{Re} z + 1$; f) $\operatorname{Re} z + \operatorname{Im} z < 1$;

g) $\alpha < \arg z < \beta$, $\gamma < \operatorname{Re} z < \delta$, where $-\frac{\pi}{2} < \alpha, \beta < \frac{\pi}{2}$, $0 < \gamma$;

h) $\operatorname{Im} \frac{z - z_1}{z - z_2} = 0$; i) $\operatorname{Re} \frac{z - z_1}{z - z_2} = 0$?

2.19. Describe the families of curves in the z-plane with equations

a) $\operatorname{Re} \frac{1}{z} = C$; b) $\operatorname{Im} \frac{1}{z} = C$; c) $\operatorname{Re} z^2 = C$; d) $\operatorname{Im} z^2 = C$,

where C is an arbitrary real number.

2.20. Describe the families of curves in the z-plane determined by the following relations:

a) $\left|\frac{z - z_1}{z - z_2}\right| = \lambda \quad (\lambda > 0)$; $\qquad \arg \frac{z - z_1}{z - z_2} = \lambda \quad (-\pi < \lambda \leqslant \pi)$.

2.21. Describe the behavior of the curve

$$|z^2 - 1| = \lambda \qquad (\lambda > 0)$$

as a function of the parameter λ.

2.22. If $\arg z$ is defined as the value of $\operatorname{Arg} z$ satisfying the inequality

$$-\pi < \operatorname{Arg} z \leqslant \pi,$$

then $\arg z$ is discontinuous on the negative real axis. On what curve is $\arg z$ discontinuous if it is defined as the value of $\operatorname{Arg} z$ satisfying the inequality[7]

a) $|z| - 2\pi < \operatorname{Arg} z \leqslant |z|$; b) $\ln |z| - 2\pi < \operatorname{Arg} z \leqslant \ln |z|$?

[7] The value $z = 0$ is excluded in every case.

2.23. Calculate the following quantities:

a) $(1 + i)^{25}$; b) $\left(\frac{1 + i\sqrt{3}}{1 - i}\right)^{30}$; c) $\left(1 - \frac{\sqrt{3} - i}{2}\right)^{24}$;

d) $\frac{(-1 + i\sqrt{3})^{15}}{(1 - i)^{20}} + \frac{(-1 - i\sqrt{3})^{15}}{(1 + i)^{20}}$.

Ans. a) $2^{12}(1 + i)$; c) $(2 - \sqrt{3})^{12}$.

2.24. Prove that

a) $(1 + i)^n = 2^{n/2}\left(\cos\frac{n\pi}{4} + i\sin\frac{n\pi}{4}\right)$;

b) $(\sqrt{3} - i)^n = 2^n\left(\cos\frac{n\pi}{6} - i\sin\frac{n\pi}{6}\right)$;

c) $(1 + \cos\alpha + i\sin\alpha)^n = 2^n\cos^n\frac{\alpha}{2}\left(\cos\frac{n\alpha}{2} + i\sin\frac{n\alpha}{2}\right)$.

2.25. Use De Moivre's theorem to express $\cos 5\Phi$, $\cos 8\Phi$, $\sin 6\Phi$ and $\sin 7\Phi$ in terms of $\cos\Phi$ and $\sin\Phi$.

2.26. Use De Moivre's theorem to express $\tan 6\Phi$ in terms of $\tan\Phi$.

2.27. Find all the values of the following roots:

a) $\sqrt[3]{1}$; b) $\sqrt[3]{i}$; c) $\sqrt[4]{-1}$; d) $\sqrt[6]{-8}$; e) $\sqrt[8]{1}$; f) $\sqrt{1 - i}$;

g) $\sqrt{3 + 4i}$; h) $\sqrt[3]{-2 + 2i}$; i) $\sqrt[5]{-4 + 3i}$; j) $\sqrt[6]{\frac{1 - i}{\sqrt{3} + i}}$;

k) $\sqrt[8]{\frac{1 + i}{\sqrt{3} - i}}$.

2.28. Prove that the sum of all the distinct nth roots of unity is zero. What geometric fact does this express?

2.29. Let m and $n \neq 0$ be any two integers. Show that $z^{m/n} = (z^{1/n})^m$ has $n/(n, m)$ distinct values, where (n, m) is the greatest common divisor of n and m. Prove that the sets of values $(z^{1/n})^m$ and $(z^m)^{1/n}$ coincide if and only if n and m are relatively prime [i.e., $(n, m) = 1$].

2.30. Calculate

a) $1 + 2\varepsilon + 3\varepsilon^2 + \cdots + n\varepsilon^{n-1}$; b) $1 + 4\varepsilon + 9\varepsilon^2 + \cdots + n^2\varepsilon^{n-1}$,

where ε is an nth root of unity.

Hint. Multiply by $1 - \varepsilon$.

Ans. a) $-\frac{n}{1 - \varepsilon}$ if $\varepsilon \neq 1$, $\frac{n(n + 1)}{2}$ if $\varepsilon = 1$.

2.31. Find all the complex numbers z satisfying the relation $\bar{z} = z^{n-1}$.

3

SETS AND FUNCTIONS. LIMITS AND CONTINUITY

7. Some Basic Definitions

From now on, we shall use geometric language when talking about complex numbers, and therefore the study of various sets of complex numbers will be equivalent to the study of various sets of points in the plane. We shall employ standard set-theoretic terminology and notation. Thus $z \in E$ means that z *belongs* to the set E (synonymously, z is a *member* of E, or z is *contained* in E), and $z \notin E$ means that z does not belong to E. By $E \subset F$ (or $F \supset E$) we mean that the set E is a *subset* of F (synonymously, E is *included* in F, or E is *contained* in F), i.e., all the points of E are points of F. Two sets E and F are said to be *equal*, and we write $E = F$, if and only if $E \subset F$, $F \subset E$. If $E \subset F$ but $E \neq F$, we say that E is a *proper subset* of F, or that E is *strictly* included in F.

By the *empty set* we mean the set containing no points at all, and we denote the empty set by the same symbol 0 as used to denote the number zero. A set E is said to be *nonempty* if $E \neq 0$. By $\{a\}$ we mean the set whose only member is a, and more generally, by $\{a_1, \ldots, a_n\}$ we mean the set whose members are $a_1, \ldots, a_n$. By the *union* of two sets E and F, written $E \cup F$, we mean the set of points belonging to at least one of the sets E and F. By the *intersection* of two sets E and F, written $E \cap F$, we mean the set of points belonging to both E and F. Two sets E and F whose intersection is empty, i.e., which are such that $E \cap F = 0$, are said to be *disjoint*. By the *complement* of a set E, written E^c, we mean the set of all points (in the complex plane)

which do not belong to E. By the *difference* $E - F$ between two sets E and F, we mean the set of points in E which do not belong to F, i.e.,

$$E - F = E \cap F^c, \qquad (E - F) \cup F = E.$$

Remark. One can also consider sets Σ whose members are themselves sets. If Σ is such a set, the union and intersection of all the sets in Σ are denoted by

$$\bigcup_{E \in \Sigma} E \quad \text{and} \quad \bigcap_{E \in \Sigma} E,$$

respectively. The sets in Σ are said to be *pairwise disjoint* (or simply *disjoint*) if $E \cap F = 0$ for any pair of disjoint sets $E \in \Sigma$, $F \in \Sigma$.

Given two nonempty sets E and $\mathscr{E}$, by a *function* f we mean a set, written

$$f = \{(z, w_z) \mid z \in E\}, \tag{3.1}$$

whose members are distinct ordered pairs of the form (z, w_z), where $z \in E$, $w_z \in \mathscr{E}$. The subscript on w_z will often be omitted, and merely serves to emphasize that w_z is one of the (possibly several) complex numbers associated with z. The set E is called the *domain* (*of definition*) of the function f, and the set $\mathscr{E}$ is called the *range* of f. If f contains no two pairs (z, w) with the same first member z, then f is said to be *single-valued*, with the *value* w at the point z. Otherwise, f is said to be *multiple-valued*, and every w such that (z, w) is contained in f is called a *value* of f at the point z. In both cases, we use $f(z)$ to denote any value of f at the point z and $\mathscr{E}_z$ to denote the set of all values of f at z. We can also write (3.1) in the abbreviated form

$$f = \{w_z \mid z \in E\},$$

or, even more simply, as

$$w = f(z) \qquad (z \in E).$$

In fact, when there is no possibility of confusion, we often omit explicit reference to the domain E. Strictly speaking, one should distinguish between the function f and its values $f(z)$ at the point z, but within the framework of this simplified notation, it is customary to use $f(z)$ to denote both the function and its values at z.

Example. The function

$$w = \operatorname{Arg} z \qquad (z \in E)$$

is a multiple-valued function defined on the set E of all nonzero points in the complex plane. In this case, $\mathscr{E}_z$ is the countably infinite set (see Problem 3.2) of points

$$\arg z + 2k\pi \qquad (k = 0, \pm 1, \pm 2, \ldots),$$

and $\mathscr{E}$ is the whole real line.

By the *inverse* f^{-1} of the function f given by (3.1) we mean the set consisting of all the ordered pairs in f *written in reverse order*, i.e., $(w, z) \in f^{-1}$ if and only if $(z, w) \in f$. Obviously, f^{-1} is itself a function with domain $\mathscr{E}$ and range E, where E is the domain and $\mathscr{E}$ the range of the original function f. Paralleling our discussion of the function f, we can write f^{-1} as

$$f^{-1} = \{(w, z_w) \mid w \in \mathscr{E}\},$$

in the abbreviated form

$$f^{-1} = \{z_w \mid w \in \mathscr{E}\},$$

or simply as

$$z = f^{-1}(w) \qquad (w \in \mathscr{E}).$$

To avoid confusion, $f^{-1}(w)$ will never be used to denote the reciprocal $1/f(w)$.

Example. The inverse of the multiple-valued function

$$w = \operatorname{Arg} z \qquad (z \in E)$$

considered in the preceding example is also multiple-valued. In fact, since Arg z is the same for all nonzero z lying on a given half-line drawn from the origin, the set of values taken by $z = \operatorname{Arg}^{-1} w$ is uncountable (see Problem 3.2).

Remark 1. Given a function $w = f(z)$ with domain E and range $\mathscr{E}$, the points of E are often called the values of the *independent variable*, and the points of $\mathscr{E}$ are often called the values of the *dependent variable*. If the domain E contains imaginary as well as real numbers, we sometimes emphasize this fact by referring to the function $w = f(z)$ as a *function of a complex variable*. Similarly, if the range $\mathscr{E}$ contains imaginary as well as real numbers, we sometimes emphasize this fact by calling $f(z)$ a *complex function*.

Remark 2. We have introduced the notion of a multiple-valued function of a complex variable in anticipation of future needs (see especially Chap. 11). However, from now on, unless the contrary is explicitly stated, all functions under discussion will be assumed to be single-valued.

Remark 3. Let $w = f(z)$, where $w = u + iv$ and $z = x + iy$, be a function of a complex variable. A special case of interest occurs when all the values of the function are real, i.e., $w = u$, $v = 0$. Then our function of a complex variable can be regarded as a real function of two real variables x and y. In fact, associating the real number u with the complex number $z = x + iy$ is equivalent to associating the real number u with the pair of real numbers x and y. In the general case where $w = u + iv$ is complex, defining a single complex function $w = f(z)$ on a set E is equivalent to defining two real functions

$$u = \operatorname{Re} w = \varphi(x, y), \qquad v = \operatorname{Im} w = \psi(x, y) \tag{3.2}$$

on E, each a function of two real variables x and y. Conversely, defining two real functions (3.2) on E is equivalent to defining a single complex function

$$w = \varphi(x, y) + i\psi(x, y) = f(z)$$

on E. Thus, it is clear that the whole theory of functions of a complex variable z can be interpreted in terms of pairs of real functions of two real variables x and y, and we shall sometimes make use of this interpretation. It should also be noted that a complex function $f(z)$ of a complex variable reduces to a single real function of two real variables whenever the range of $f(z)$ is a subset of the real line; in fact, $f(z)$ reduces to a single real function of a single real variable whenever the range of $f(z)$ is a subset of the real line and the domain of $f(z)$ is a line parallel to either the real or imaginary axis.

Example 1. The single complex function

$$w = z^2 = (x + iy)^2 = x^2 - y^2 + 2ixy$$

corresponds to the two real functions

$$u = x^2 - y^2, \qquad v = 2xy.$$

Example 2. The two real functions

$$u = x^2 - y^2, \qquad v = e^{x^2 + 3y^2}$$

correspond to the single complex function

$$w = u + iv = (x^2 - y^2) + ie^{x^2 + 3y^2} = f(z).$$

If $f(z)$ is a single-valued function with domain E and range $\mathscr{E}$, the point $w = f(z) \in \mathscr{E}$ is said to be the *image* of the point $z \in E$ [under $f(z)$]. Similarly, if $D \subset E$, the set of all points $w = f(z)$ such that $z \in D$ is called the *image* of D, denoted by $f(D)$. Let $\mathscr{D}$ be any subset of $\mathscr{E}$ containing $f(D)$. Then we say that $f(z)$ is a *mapping* of D *into* $\mathscr{D}$. If $\mathscr{D} = f(D)$ and we want to emphasize this fact, we say that $f(z)$ is a *mapping* of D *onto* $\mathscr{D}$. Thus, every "onto" mapping is an "into" mapping, but not conversely. A mapping of a set into itself is often called a *transformation*.

If $f(z)$ is single-valued, the inverse function $f^{-1}(w)$ will in general be multiple-valued. If $f^{-1}(w)$ is single-valued, the original function or mapping $f(z)$ is said to be *one-to-one*. Then the point $z \in E$ which is the image of the point $w \in \mathscr{E}$ under $f^{-1}(w)$ is called the *inverse image* or *pre-image* of w [under $f(z)$]. Similarly, if $\mathscr{D} \subset \mathscr{E}$, the set of all points $z = f^{-1}(w)$ such that $w \in \mathscr{D}$ is called the *inverse image* or *pre-image* of $\mathscr{D}$, denoted by $f^{-1}(\mathscr{D})$. A one-to-one function $f(z)$ is said to establish a *one-to-one correspondence*, indicated by $z \leftrightarrow w$, between the points z in its domain and the points w in its range.

Remark. In the case of multiple-valued functions, images and inverse images are defined in the natural way. For example, if $f(z)$ takes several

values at a given point $z \in E$, each of these values is called an *image* of the point z. Similarly, if there are several points $z \in E$ such that $f(z) = w$, where w is a given point of $\mathcal{E}$, then each of the points z is called an *inverse image* of w. In terms of the definition (3.1) of a function, the images of a given point $z \in E$ are all those points w appearing in some ordered pair (z, w) with z as its first member, and the inverse images of a given point $w \in \mathcal{E}$ are all those points z appearing in some ordered pair (z, w) with w as its second member.

8. Sequences of Complex Numbers. Limit Points and Limits of Sequences

Let I be the set of all positive integers $1, 2, \ldots, n, \ldots$. Then any single-valued function $f(z)$ with domain I is called an *infinite sequence*, or simply a *sequence*, and the values $z_n = f(n)$ are called the terms of the sequence. The sequence itself is usually denoted by listing its terms in order of increasing indices, i.e.,

$$z_1, z_2, \ldots, z_n, \ldots,$$

or by the more concise notation $\{z_n\}$ (not to be confused with the set whose only member is z_n). If $k_1, k_2, \ldots, k_n, \ldots$ is a sequence of positive integers such that $k_n < k_{n+1}$ for all $n = 1, 2, \ldots$, then the corresponding terms of the sequence $\{z_n\}$, i.e., $z_{k_1}, z_{k_2}, \ldots, z_{k_n}, \ldots$ form a new sequence $\{z_{k_n}\}$, called a *subsequence* of the sequence $\{z_n\}$. For example, the sequences

$$z_1, z_3, z_5, \ldots, z_{2n-1}, \ldots,$$
$$z_2, z_4, z_6, \ldots, z_{2n}, \ldots,$$
$$z_1, z_4, z_9, \ldots, z_{n^2}, \ldots$$

are all subsequences of the sequence $\{z_n\}$.

Remark. The subsequences of $\{z_n\}$ can also be written in the form

$$z_{n_1}, z_{n_2}, \ldots, z_{n_k}, \ldots,$$

or briefly $\{z_{n_k}\}$, where $n_1, n_2, \ldots, n_k, \ldots$ is a sequence of positive integers, such that $n_k < n_{k+1}$ for all $k = 1, 2, \ldots$, so that k rather than n runs through the positive integers. In other words, the running index is always the "lowest subscript," regardless of the particular symbol used.

Let z be a point of the complex plane. Then the interior of the circle of radius ε with z as its center, i.e., the set of all points z' satisfying the inequality $|z' - z| < \varepsilon$, is called an *ε-neighborhood* of z, or simply a *neighborhood* of z, and is denoted by $\mathcal{N}(z; \varepsilon)$, or simply by $\mathcal{N}(z)$ if ε is unimportant.[1]

[1] A neighborhood of z can also be defined as the interior of *any* circle containing z, but there is essentially no loss of generality in assuming that z is the center of the circle.

A point ζ is said to be a *limit point* of the sequence $\{z_n\}$ if for every neighborhood $\mathcal{N}(\zeta)$ of the point ζ, there is a subsequence $\{z_{k_n}\}$ all of whose terms belong to $\mathcal{N}(\zeta)$. Equivalently, ζ is said to be a limit point of $\{z_n\}$ if every neighborhood $\mathcal{N}(\zeta)$ contains infinitely many terms of $\{z_n\}$. It is important to note that when referring to a sequence, the phrase "infinitely many terms" means infinitely many terms with distinct indices, but only finitely many of these terms need be distinct complex numbers.

Example 1. The point 0 is a limit point of the sequence

$$1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots$$

Example 2. The points 0 and 1 are limit points of the sequence

$$0, 1, 0, 1, 0, 1, \ldots \tag{3.3}$$

For the point 0, the subsequence $0, 0, 0, \ldots$ figuring in the definition of a limit point consists of the terms of (3.3) with odd indices, while for the point 1, the appropriate subsequence $1, 1, 1, \ldots$ consists of the terms with even indices.

We note that if $\{z_{l_n}\}$ is a subsequence of the sequence $\{z_{k_n}\}$, and if $\{z_{k_n}\}$ is a subsequence of the sequence $\{z_n\}$, then $\{z_{l_n}\}$ is a subsequence of the sequence $\{z_n\}$. It follows that any limit point of a subsequence is a limit point of the original sequence. The converse need not be true, as shown by the sequence (3.3), where the subsequence $0, 0, 0, \ldots$ has only one limit point, whereas the original sequence has two limit points.

A sequence $\{z_n\}$ is said to be *bounded* if there exists a neighborhood of the point $z = 0$ containing all the terms of the sequence, i.e., if there exists a number $\rho > 0$ such that $|z_n| < \rho$ for all $n = 1, 2, \ldots$.

THEOREM 3.1 (*Bolzano-Weierstrass principle for sequences*). *Every bounded sequence $\{z_n\}$ has at least one limit point.*

Proof. Let Q_0 be a square with sides parallel to the coordinate axes and center at the origin,[2] which contains all the terms of the sequence $\{z_n\}$. The coordinate axes divide Q_0 into four squares, at least one of which, say Q_1, contains infinitely many terms of $\{z_n\}$. Similarly, dividing Q_1 into four congruent squares, we can find a new square, say Q_2, containing infinitely many terms of $\{z_n\}$ (see Figure 3.1). Continuing this construction, we arrive at a sequence of squares

$$Q_0 \supset Q_1 \supset Q_2 \supset \cdots \supset Q_n \supset \cdots,$$

[2] Here, by a *square* we mean a two-dimensional closed interval of the form $a - c \leqslant x \leqslant a + c$, $b - c \leqslant y \leqslant b + c$ $(c > 0)$. The point (a, b) is called the *center* of the square.

each strictly including the next and each containing infinitely many terms of $\{z_n\}$. The projections of these squares on the x and y-axes, respectively, form two "nested" sequences of closed intervals

$$X_0 \supset X_1 \supset X_2 \supset \cdots \supset X_n \supset \cdots,$$
$$Y_0 \supset Y_1 \supset Y_2 \supset \cdots \supset Y_n \supset \cdots,$$

where each of the intervals X_n is twice as long as the next, and similarly for the intervals Y_n. As is familiar from elementary analysis,[3] there is a unique point ξ in all the intervals X_n and a unique point η in all the intervals Y_n. Therefore, there is a unique point $\zeta = \xi + i\eta$ in all the squares Q_n ($n = 0, 1, 2, \ldots$).

We now show that ζ is a limit point of the sequence $\{z_n\}$. Given any neighborhood $\mathcal{N}(\zeta; \varepsilon)$ of the point ζ, we can choose n so large that the square Q_n belongs entirely to $\mathcal{N}(\zeta; \varepsilon)$. This can be done by requiring that the length of the diagonal of the square Q_n be less than ε. Since the length of the diagonal of Q_n is obviously $l/2^n$, where l is the length of the diagonal of the square Q_0, this requirement is satisfied if $l/2^n < \varepsilon$, and then, since Q_n contains infinitely many terms of $\{z_n\}$, the same is true of $\mathcal{N}(\zeta; \varepsilon)$. Therefore ζ is a limit point of $\{z_n\}$, and the theorem is proved.

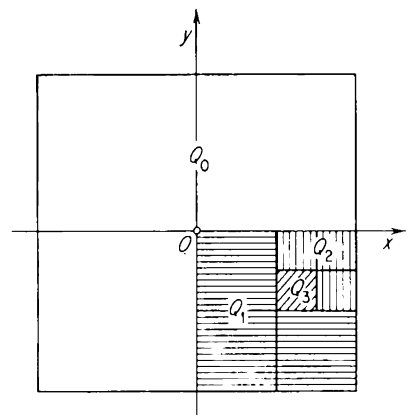

FIGURE 3.1

A bounded sequence $\{z_n\}$ which has only one limit point ζ is said to be *convergent*, and ζ is called its *limit*. We also say that $\{z_n\}$ *converges* to ζ, a fact indicated by writing $z_n \to \zeta$ as $n \to \infty$, or

$$\lim_{n \to \infty} z_n = \zeta.$$

It is clear that every subsequence $\{z_{k_n}\}$ of a convergent sequence $\{z_n\}$ with limit ζ converges to the same limit ζ. In fact, $\{z_{k_n}\}$ is bounded, since the whole sequence $\{z_n\}$ is bounded, and therefore, according to Theorem 3.1, $\{z_{k_n}\}$ has at least one limit point. But the limit points of $\{z_{k_n}\}$ are simultaneously limit points of $\{z_n\}$, so that $\{z_{k_n}\}$ has only one limit point ζ, as asserted.

THEOREM 3.2. *A sequence $\{z_n\}$ converges to the limit ζ if and only if given any $\varepsilon > 0$, there exists an integer $N(\varepsilon) > 0$ such that $|z_n - \zeta| < \varepsilon$ for all $n > N(\varepsilon)$, i.e., if and only if any given neighborhood of ζ contains all the terms of $\{z_n\}$ starting from some value of n.*

[3] See e.g., K. Knopf, *Infinite Sequences and Series* (translated by F. Bagemihl), Dover Publications, Inc., New York (1956), p. 39.

Proof. First we prove that the condition is necessary. Suppose that $\{z_n\}$ converges to ζ, and suppose that for some $\varepsilon > 0$, infinitely many terms

$$z_{k_1}, z_{k_2}, \ldots, z_{k_n}, \ldots \qquad (k_1 < k_2 < \cdots < k_n < \cdots)$$

of the sequence $\{z_n\}$ belong to the complement of $\mathcal{N}(\zeta; \varepsilon)$. Then these terms themselves form a sequence $\{z_{k_n}\}$ which, like the original sequence $\{z_n\}$, is bounded. Therefore, according to Theorem 3.1, $\{z_{k_n}\}$ must have at least one limit point ζ'. But $\zeta' \neq \zeta$, since $\mathcal{N}(\zeta; \varepsilon)$ does not contain any points of $\{z_{k_n}\}$, while, on the other hand, ζ' must be a limit point of $\{z_n\}$, since it is a limit point of $\{z_{k_n}\}$. Thus, we have found a limit point of $\{z_n\}$ different from its (unique) limit point ζ. This contradiction proves that for any $\varepsilon > 0$, only finitely many terms of $\{z_n\}$ belong to the complement of $\mathcal{N}(\zeta; \varepsilon)$, i.e., all the terms of $\{z_n\}$ starting from some value of n belong to $\mathcal{N}(\zeta; \varepsilon)$.

Next we prove that the condition is sufficient. If the condition holds, the sequence $\{z_n\}$ is bounded, since all the points of $\{z_n\}$ starting from some value of n belong to the neighborhood $\mathcal{N}(\zeta; 1)$, say, and we can certainly find a neighborhood of the point $z = 0$ with large enough radius to contain both $\mathcal{N}(\zeta; 1)$ and the finite number of points of $\{z_n\}$ which do not belong to $\mathcal{N}(\zeta; 1)$. It follows from Theorem 3.1 that the sequence $\{z_n\}$ must have at least one limit point. But no point ζ' different from ζ can be a limit point of $\{z_n\}$. To see this, we choose $\varepsilon < \frac{1}{2}|\zeta' - \zeta|$ and note that only finitely many terms of the sequence $\{z_n\}$ belong to the complement of $\mathcal{N}(\zeta; \varepsilon)$. Therefore, only finitely many terms of $\{z_n\}$ belong to $\mathcal{N}(\zeta'; \varepsilon)$, which means that ζ' is not a limit point of $\{z_n\}$. Thus ζ is the unique limit point of $\{z_n\}$, i.e., $\{z_n\}$ converges to ζ, and the theorem is proved.

THEOREM 3.3. *Every bounded sequence $\{z_n\}$ has a convergent subsequence.*

Proof. Let ζ be a limit point of the sequence $\{z_n\}$ (according to Theorem 3.1, such a point exists). Then any neighborhood of ζ contains infinitely many terms of $\{z_n\}$ (recall that only finitely many of these terms need be distinct complex numbers). First consider the neighborhood $\mathcal{N}(\zeta; 1)$, and let z_{k_1} be any term of $\{z_n\}$ belonging to $\mathcal{N}(\zeta; 1)$. Next consider the neighborhood $\mathcal{N}(\zeta; \frac{1}{2})$, and from the infinitely many terms of $\{z_n\}$ in $\mathcal{N}(\zeta; \frac{1}{2})$ choose a term z_{k_2} with index greater than k_1. Suppose repeating this procedure n times leads to n terms $z_{k_1}, z_{k_2}, \ldots, z_{k_n}$ of $\{z_n\}$, where $k_1 < k_2 < \cdots < k_n$, such that z_k belongs to the neighborhood $\mathcal{N}(\zeta; 1/k)$, $k = 1, 2, \ldots, n$. Then, since the neighborhood $\mathcal{N}(\zeta; 1/(n+1))$ con-

tains infinitely many terms of $\{z_n\}$, one of these terms, say $z_{k_{n+1}}$, must have an index greater than k_n. In this way, we obtain a subsequence $\{z_{k_n}\}$ of the sequence $\{z_n\}$ such that

$$|z_{k_n} - \zeta| < \frac{1}{n} \qquad (n = 1, 2, \ldots).$$

Obviously, all the terms z_{k_n} of this subsequence such that $n > 1/\varepsilon$ belong to the neighborhood $\mathcal{N}(\zeta; \varepsilon)$. In other words, $\{z_{k_n}\}$ converges to ζ, i.e., $\{z_n\}$ has a convergent subsequence, as asserted.

The next theorem gives a criterion for the convergence of a sequence which does not require prior knowledge of the limit of the sequence.

THEOREM 3.4 (*Cauchy convergence criterion*). *The sequence* $\{z_n\}$ *is convergent if and only if given any* $\varepsilon > 0$, *there exists an integer* $N(\varepsilon) > 0$ *such that*

$$|z_{n+p} - z_n| < \varepsilon \tag{3.4}$$

for all $n > N(\varepsilon)$ *and* $p > 0$.

Proof. First we prove that the condition (3.4) is necessary. If $\{z_n\}$ converges to a limit ζ, then

$$|z_n - \zeta| < \frac{\varepsilon}{2}$$

for all n greater than a certain integer we call $N(\varepsilon)$. It follows that

$$|z_{n+p} - \zeta| < \frac{\varepsilon}{2},$$

and hence

$$|z_{n+p} - z_n| = |(z_{n+p} - \zeta) + (\zeta - z_n)| \leqslant |z_{n+p} - \zeta| + |z_n - \zeta| < \varepsilon$$

for all $n > N(\varepsilon)$ and (all) $p > 0$.

Next we prove that the condition (3.4) is sufficient. Choosing $\varepsilon = 1$, we have

$$|z_{n+p} - z_n| < 1,$$

starting from a certain value of n. Fixing one of these values of n, say n_0, we find that

$$|z_{n_0+p} - z_{n_0}| < 1,$$

so that all the points $z_{n_0+1}, z_{n_0+2}, \ldots$ belong to the neighborhood $\mathcal{N}(z_{n_0}; 1)$, i.e., lie inside the circle with center z_{n_0} and radius 1. Clearly, there is a neighborhood of the point $z = 0$ with large enough radius to contain both $\mathcal{N}(z_{n_0}; 1)$ and the finite number of points $z_1, z_2, \ldots, z_{n_0}$.

Thus the sequence $\{z_n\}$ is bounded, and hence, according to Theorem 3.1, $\{z_n\}$ has at least one limit point.

It remains to show that $\{z_n\}$ cannot have two different limit points. Suppose the opposite is true, i.e., let ζ and ζ' ($\zeta' \neq \zeta$) be two limit points of the sequence $\{z_n\}$. Choosing $\varepsilon = \frac{1}{3}|\zeta - \zeta'|$, we find that the inequality

$$|z_{n+p} - z_n| < \tfrac{1}{3}|\zeta - \zeta'| \tag{3.5}$$

holds for all sufficiently large n, say $n > N$. On the other hand, each of the neighborhoods $\mathcal{N}(\zeta; \varepsilon)$ and $\mathcal{N}(\zeta'; \varepsilon)$ must contain infinitely many

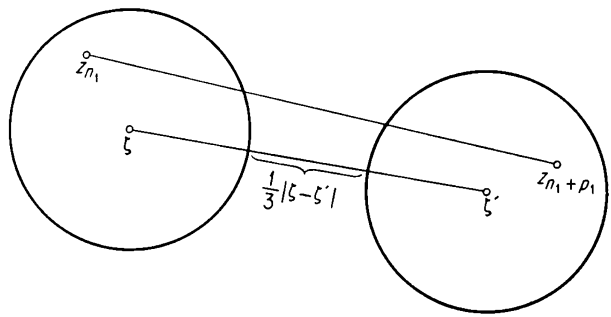

FIGURE 3.2

terms of $\{z_n\}$. Let z_{n_1} be a term with index larger than N belonging to $\mathcal{N}(\zeta; \varepsilon)$, and let $z_{n_1+p_1}$ be a term belonging to $\mathcal{N}(\zeta'; \varepsilon)$. Then it is obvious from Figure 3.2 that

$$|z_{n_1+p_1} - z_{n_1}| > \tfrac{1}{3}|\zeta - \zeta'|,$$

which contradicts the inequality (3.5). This contradiction implies that $\{z_n\}$ has only one limit point and hence converges.

9. Convergence of the Real and Imaginary Parts, Moduli and Arguments of a Complex Sequence

Given a complex sequence $\{z_n\}$, let $z_n = x_n + iy_n$, i.e., let z_n and y_n be the real and imaginary parts of the term z_n. Then every sequence of complex numbers $\{z_n\}$ corresponds to two sequences of real numbers $\{x_n\}$ and $\{y_n\}$.

THEOREM 3.5. *The complex sequence $\{z_n\}$, where $z_n = x_n + iy_n$, converges to the limit $\zeta = \xi + i\eta$ if and only if the real sequences $\{x_n\}$ and $\{y_n\}$ converge to the limits ξ and η, respectively.*

Proof. If $z_n \to \zeta$ as $n \to \infty$, then for any $\varepsilon > 0$ there exists an integer $N(\varepsilon) > 0$ such that $|z_n - \zeta| < \varepsilon$ if $n > N(\varepsilon)$. It follows that

$$|x_n - \xi| = |\mathrm{Re}\,(z_n - \zeta)| \leqslant |z_n - \zeta| < \varepsilon,$$
$$|y_n - \eta| = |\mathrm{Im}\,(z_n - \zeta)| \leqslant |z_n - \zeta| < \varepsilon$$

if $n > N(\varepsilon)$, and hence $x_n \to \xi$, $y_n \to \eta$ as $n \to \infty$.

Conversely, if $x_n \to \xi$, $y_n \to \eta$ as $n \to \infty$, then for any $\varepsilon > 0$ there exists an integer $N(\varepsilon) > 0$ such that $|x_n - \xi| < \varepsilon$, $|y_n - \eta| < \varepsilon$ if $n > N(\varepsilon)$. It follows that

$$|z_n - \zeta| = |(x_n - \xi) + i(y_n - \eta)| \leqslant |x_n - \xi| + |y_n - \eta| < 2\varepsilon$$

if $n > N(\varepsilon)$, and hence $z_n \to \zeta$ as $n \to \infty$, since ε is arbitrary.

Remark. According to Theorem 3.5, convergence of a complex sequence is equivalent to the convergence of two real sequences, i.e., those formed from the real and imaginary parts of the complex sequence. This fact can be used to *define* convergence of a complex sequence in terms of the convergence of its real and imaginary parts (this is done in the obvious way). Then the theorems of Sec. 8 become simple consequences of the corresponding theorems for real sequences, which we have actually rederived for the case of complex numbers. Moreover, by using Theorem 3.4, the familiar results concerning algebraic operations on convergent real sequences can be carried over without change to the case of complex sequences. Thus, given two complex sequences $\{z_n\}$ and $\{z'_n\}$, suppose that $z_n \to \zeta$, $z'_n \to \zeta'$ as $n \to \infty$. Then

$$z_n \pm z'_n \to \zeta \pm \zeta', \qquad z_n z'_n \to \zeta\zeta'$$

and

$$\frac{z_n}{z'_n} \to \frac{\zeta}{\zeta'} \tag{3.6}$$

as $n \to \infty$, where in writing (3.6) it is assumed that $z'_n \neq 0$ $(n = 1, 2, \ldots)$ and $\zeta' \neq 0$.

To decide whether or not a given complex sequence $\{z_n\}$ converges, we can examine the sequences of moduli and arguments $\{|z_n|\}$, $\{\arg z_n\}$ instead of the sequences of real and imaginary parts $\{x_n\}$, $\{y_n\}$. In doing so, we encounter various possibilities:

1. A necessary and sufficient condition for a complex sequence $\{z_n\}$ to converge to *zero* is that the sequence of moduli $\{|z_n|\}$ converge to zero. (The sequence of arguments $\{\arg z_n\}$ need not be considered at all, and in fact $\{\arg z_n\}$ may not converge.) To see this, it is enough to note that

$$|z_n - 0| = |z_n| < \varepsilon \qquad \text{for all } n > N(\varepsilon)$$

is a necessary and sufficient condition not only for the complex sequence $\{z_n\}$ to converge to zero, but also for the real sequence $\{|z_n|\}$ to converge to zero.

2. In the general case, simultaneous convergence of the sequences $\{|z_n|\}$ and $\{\arg z_n\}$ is sufficient for convergence of the sequence $\{z_n\}$, i.e., if

$$\lim_{n\to\infty} |z_n| = r, \qquad \lim_{n\to\infty} \arg z_n = \varphi,$$

then

$$\lim_{n\to\infty} z_n = r(\cos\varphi + i\sin\varphi).$$

To see this, we note that

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} \operatorname{Re} z_n = \lim_{n\to\infty} |z_n| \cos(\arg z_n) = r\cos\varphi,$$
$$\lim_{n\to\infty} y_n = \lim_{n\to\infty} \operatorname{Im} z_n = \lim_{n\to\infty} |z_n| \sin(\arg z_n) = r\sin\varphi.$$

The result now follows from Theorem 3.5.

3. Conversely, if the sequence $\{z_n\}$ converges, the sequence of moduli $\{|z_n|\}$ also converges. In fact, if

$$\lim_{n\to\infty} x_n = \zeta, \qquad \lim_{n\to\infty} y_n = \eta,$$

then

$$\lim_{n\to\infty} |z_n| = \lim_{n\to\infty} \sqrt{x_n^2 + y_n^2} = \sqrt{\xi^2 + \eta^2}.$$

4. The sequence of arguments $\{\arg z_n\}$ of a convergent sequence need not converge even in the case where

$$\lim_{n\to\infty} z_n \neq 0.$$

For example, let

$$z_n = -1 + (-1)^n \frac{i}{n}.$$

Then $z_n \to -1$ as $n \to \infty$, but the sequence $\{\arg z_n\}$ does not converge, since

$$\arg z_{2k} = \pi - \arctan\frac{1}{2k}, \qquad \arg z_{2k+1} = -\pi + \arctan\frac{1}{2k+1}.$$

However, in this case we can find a convergent sequence of values of $\operatorname{Arg} z_n$. In fact, let Φ_n denote the unique value of $\operatorname{Arg} z_n$ satisfying the inequality

$$\frac{\pi}{2} < \Phi_n < \frac{3\pi}{2}.$$

Then the sequence $\{\Phi_n\}$ converges to the limit π, since

$$\Phi_{2k} = \pi - \arctan\frac{1}{2k}, \qquad \Phi_{2k+1} = \pi + \arctan\frac{1}{2k+1}.$$

5. The argument used in the last example can be made perfectly general. Suppose that

$$\lim_{n \to \infty} z_n = \zeta \neq 0,$$

and let Φ be any value of Arg ζ. Let n_0 be the smallest value of n for which there exists a neighborhood $\mathcal{N}(\zeta)$ containing all the points $z_n, z_{n+1}, \ldots$ *but not the origin* (see Figure 3.3), and for $n \geqslant n_0$ let Φ_n denote the unique value of Arg z_n satisfying the inequality

$$|\Phi_n - \Phi| < \frac{\pi}{2}.$$

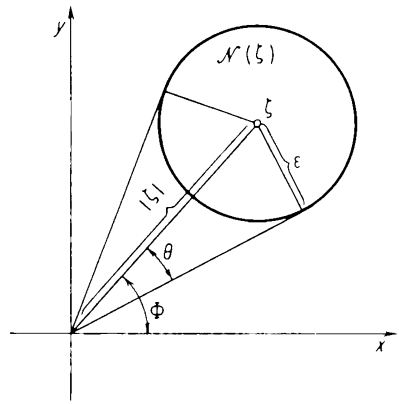

FIGURE 3.3

Then $\Phi_n \to \Phi$ as $n \to \infty$. To see this, let θ be any number such that $0 < \theta < \pi/2$, and let $\varepsilon = |\zeta| \sin \theta$. It is obvious that $0 < \varepsilon < |\zeta|$. Since $z_n \to \zeta$, we can always find an integer $N(\varepsilon) > 0$ such that

$$|z_n - \zeta| < \varepsilon$$

if $n > N(\varepsilon)$. Every point z_n such that $n > N(\varepsilon)$ belongs to the neighborhood $\mathcal{N}(\zeta)$, and hence has a uniquely defined value of $\Phi_n = \text{Arg } z_n$. Moreover, if $n > N(\varepsilon) \equiv N'(\theta)$, then

$$|\Phi_n - \Phi| < \arcsin \frac{\varepsilon}{|\zeta|} = \theta$$

(see Figure 3.3), and hence

$$\lim_{n \to \infty} \Phi_n = \Phi,$$

as asserted. In other words, if the sequence $\{z_n\}$ converges to a limit $\zeta \neq 0$, then, given any value of $\Phi = \text{Arg } \zeta$, there exists a sequence of values of Arg z_n which converges to Φ, a fact indicated by writing

$$\lim_{n \to \infty} \text{Arg } z_n = \text{Arg } \zeta.$$

(Note that the expressions Arg $z_1, \ldots,$ Arg $z_{n_0 - 1}$ are multiple-valued or even meaningless, since there is nothing to prevent the numbers $z_1, \ldots, z_{n_0 - 1}$ from vanishing.)

10. Series with Complex Terms

Let

$$\sum_{n=1}^{\infty} z_n = z_1 + \cdots + z_n + \cdots \tag{3.7}$$

be an *infinite series* (or simply a *series*) with complex terms z_n, and let

$$\{s_n\} = \{z_1 + \cdots + z_n\}, \tag{3.8}$$

be the sequence of *partial sums* of (3.7). Then the series (3.7) is said to *converge* if the sequence (3.8) converges, and the limit of the sequence is called the *sum* of the series. Moreover, if

$$\lim_{n \to \infty} s_n = s,$$

we write

$$\sum_{n=1}^{\infty} z_n = z_1 + \cdots + z_n + \cdots = s.$$

A series which does not converge is said to be *divergent*.

If we set

$$z_n = x_n + iy_n, \qquad s_n = \sigma_n + i\tau_n = (x_1 + \cdots + x_n) + i(y_1 + \cdots + y_n),$$

it follows from Theorem 3.5 that the series (3.7) converges if and only if both of the sequences $\{\sigma_n\}$ and $\{\tau_n\}$ converge. But these are the sequences of partial sums of the following series with real terms

$$\sum_{n=1}^{\infty} x_n = x_1 + \cdots + x_n + \cdots, \tag{3.9}$$

$$\sum_{n=1}^{\infty} y_n = y_1 + \cdots + y_n + \cdots. \tag{3.10}$$

Therefore, a series with complex terms converges if and only if the series formed from the real and imaginary parts of its terms converge. Moreover, if

$$\sum_{n=1}^{\infty} x_n = \lim_{n \to \infty} (x_1 + \cdots + x_n) = \sigma,$$

$$\sum_{n=1}^{\infty} y_n = \lim_{n \to \infty} (y_1 + \cdots + y_n) = \tau,$$

then

$$s = \lim_{n \to \infty} (z_1 + \cdots + z_n) = \sigma + i\tau.$$

Thus the sum of the series (3.7) equals $\sigma + i\tau$, where σ and τ are the sums of the series (3.9) and (3.10), and conversely, the sums of (3.9) and (3.10) are the real and imaginary parts of the sum of (3.7).

Applying the Cauchy convergence criterion (Theorem 3.4) to the sequence of partial sums (3.8), and noting that

$$s_{n+p} - s_n = z_{n+1} + \cdots + z_{n+p},$$

we find that the series (3.7) converges if and only if given any $\varepsilon > 0$, there exists an integer $N(\varepsilon) > 0$ such that

$$|z_{n+1} + \cdots + z_{n+p}| < \varepsilon$$

for all $n > N(\epsilon)$ and $p > 0$. In particular, a necessary condition for the series (3.7) to converge is that

$$\lim_{n \to \infty} z_n = 0.$$

The series (3.7) is said to be *absolutely convergent* if the series

$$\sum_{n=1}^{\infty} |z_n| = |z_1| + \cdots + |z_n| + \cdots \tag{3.11}$$

formed from the moduli of its terms converges. Since

$$|z_{n+1} + \cdots + z_{n+p}| \leqslant |z_{n+1}| + \cdots + |z_{n+p}|,$$

a series with complex terms is convergent if it is absolutely convergent. The converse is not true, as shown by the familiar example

$$1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \cdots$$

It follows from the inequalities

$$|x_n| \leqslant |z_n|, \quad |y_n| \leqslant |z_n| \quad \text{and} \quad |z_n| \leqslant |x_n| + |y_n|$$

that the series (3.7) is absolutely convergent if and only if the series (3.9) and (3.10) are absolutely convergent.

Let the series (3.7) be absolutely convergent with sum $\sigma + i\tau$. Then the series (3.9) and (3.10) are also absolutely convergent with sums σ and τ, respectively. Suppose we make an arbitrary rearrangement of the terms of (3.7), thereby obtaining a new series

$$\sum_{n=1}^{\infty} z_{k_n} = z_{k_1} + \cdots + z_{k_n} + \cdots, \tag{3.12}$$

where the sequence $k_1, \ldots, k_n, \ldots$ contains all the positive integers, each just once. Correspondingly, the series formed from the real and imaginary parts of the series (3.12) have the form

$$\sum_{n=1}^{\infty} x_{k_n} = x_{k_1} + \cdots + x_{k_n} + \cdots,$$

$$\sum_{n=1}^{\infty} y_{k_n} = y_{k_1} + \cdots + y_{k_n} + \cdots.$$

These series are obtained by rearranging the terms of the absolutely convergent series (3.9) and (3.10), and therefore, as is well known,[4] they converge to their previous sums σ and τ. It follows that the series (3.12) with complex terms also converges to its original sum $\sigma + i\tau$.

Since absolute convergence of the series (3.7) is equivalent to convergence of the series (3.11), to test for absolute convergence we can use any known convergence test for series with nonnegative terms. In particular, we note

[4] See e.g., R. C. Buck, *op. cit.*, Theorem 11, p. 115.

that a sufficient condition for the absolute convergence of the series (3.7) is that starting from some value of n, the inequality

$$|z_{n+1}| \leqslant q|z_n| \qquad (0 < q < 1) \tag{3.13}$$

(the *ratio test*), or the inequality

$$\sqrt[n]{|z_n|} \leqslant q \qquad (0 < q < 1) \tag{3.14}$$

(the *root test*) should hold.[5] The root test is stronger than the ratio test in the sense that (3.14) is implied by (3.13), but not conversely.

The following sufficient condition for convergence is often useful:

THEOREM 3.6. *Let* $\{a_n\}$ *and* $\{b_n\}$ *be two complex sequences such that*

1. *The partial sums of the series*

$$\sum_{n=1}^{\infty} a_n$$

form a bounded sequence, i.e.,

$$|a_1 + \cdots + a_n| \leqslant M \qquad (n = 1, 2, \ldots),$$

for some constant $M > 0$;

2. *The series*

$$\sum_{n=1}^{\infty} |b_n - b_{n+1}|$$

converges, and $b_n \to 0$ *as* $n \to \infty$.

Then the series

$$\sum_{n=1}^{\infty} a_n b_n \tag{3.15}$$

also converges.

Proof. If we denote the partial sum

$$\sum_{n=1}^{\infty} a_k$$

by α_n, then $a_n = \alpha_n - \alpha_{n-1}$ and[6]

$$\begin{aligned} \sum_{k=n+1}^{n+p} a_k b_k &= \sum_{k=n+1}^{n+p} (\alpha_k - \alpha_{k-1}) b_k = \sum_{k=n+1}^{n+p} \alpha_k b_k - \sum_{k=n+1}^{n+p} \alpha_{k-1} b_k \\ &= \sum_{k=n+1}^{n+p} \alpha_k b_k - \sum_{k=n}^{n+p-1} \alpha_k b_{k+1} \\ &= \alpha_{n+p} b_{n+p} - \alpha_n b_{n+1} - \sum_{k=n+1}^{n+p-1} \alpha_k (b_{k+1} - b_k). \end{aligned} \tag{3.16}$$

[5] See e.g., R. C. Buck, *op. cit.*, Theorem 5, p. 108, and Theorem 6, p. 109.

[6] The transformation leading from the left-hand side to the right-hand side of (3.16) is called the *Abel transformation*, and is the analogue of integration by parts.

It follows from (3.16) that

$$\left|\sum_{k=n+1}^{n+p} a_k b_k\right| \leqslant |\alpha_{n+p}|\,|b_{n+p}| + |\alpha_n|\,|b_{n+1}| + \sum_{k=n+1}^{n+p-1} |\alpha_k|\,|b_k - b_{k+1}|$$
$$\leqslant M\left(|b_{n+p}| + |b_{n+1}| + \sum_{k=n+1}^{n+p-1} |b_k - b_{k+1}|\right).$$

By hypothesis, there exists an integer $N(\varepsilon) > 0$ such that

$$|b_n| < \frac{\varepsilon}{3M}, \qquad \sum_{k=n+1}^{n+p} |b_k - b_{k+1}| < \frac{\varepsilon}{3M}$$

for all $n > N(\varepsilon)$ and $p > 0$, and hence

$$\left|\sum_{k=n+1}^{n+p} a_k b_k\right| < \varepsilon.$$

Therefore the series (3.15) converges, as asserted.

The familiar operations involving series with real terms carry over to series with complex terms. In particular, we note the following results, all of which are easily proved:

1. For any integer $n > 1$, the two series

$$\sum_{n=1}^{\infty} z_n, \qquad \sum_{n=N}^{\infty} z_n$$

converge or diverge together.

2. If the series $z_1 + \cdots + z_n + \cdots$ converges and its sum is s, then the series $\lambda z_1 + \cdots + \lambda z_n + \cdots$ converges and its sum is λs.

3. If the series $z_1 + \cdots + z_n + \cdots$ converges and its sum is s, then

$$(z_1 + \cdots + z_{k_1}) + (z_{k_1+1} + \cdots + z_{k_2}) + \cdots + (z_{k_n+1} + \cdots + z_{k_{n+1}}) + \cdots = s, \tag{3.17}$$

where $0 < k_1 < k_2 < \cdots < k_n < k_{n+1} < \cdots$.

4. If

$$\sum_{n=1}^{\infty} z_n = s, \qquad \sum_{n=1}^{\infty} z'_n = s',$$

then

$$\sum_{n=1}^{\infty} (z_n \pm z'_n) = s \pm s'.$$

THEOREM 3.7. *If the series*

$$\sum_{n=1}^{\infty} z_n \quad \textit{and} \quad \sum_{n=1}^{\infty} z'_n \tag{3.18}$$

are absolutely convergent, with sums s and s', respectively, then the series

$$\sum_{n=1}^{\infty} (z_1 z'_n + z_2 z'_{n-1} + \cdots + z_n z'_1)$$

is also absolutely convergent, with sum ss'.

Proof. The series

$$\begin{aligned} &|z_1|\,|z'_1| + |z_1|\,|z'_2| + |z_2|\,|z'_1| + \cdots \\ &\qquad + |z_1|\,|z'_n| + |z_2|\,|z'_{n-1}| + \cdots + |z_n|\,|z'_1| + \cdots \\ &= |z_1 z'_1| + |z_1 z'_2| + |z_2 z'_1| + \cdots \\ &\qquad + |z_1 z'_n| + |z_2 z'_{n-1}| + \cdots + |z_n z'_1| + \cdots \end{aligned} \tag{3.19}$$

converges,[7] being the product of the two absolutely convergent real series

$$\sum_{n=1}^{\infty} |z_n|, \quad \sum_{n=1}^{\infty} |z'_n|.$$

Therefore the complex series

$$z_1 z'_1 + z_1 z'_2 + z_2 z'_1 + \cdots + z_1 z'_n + z_2 z'_{n-1} + \cdots + z_n z'_1 + \cdots \tag{3.20}$$

is absolutely convergent. Consequently, the sum of the series (3.20) does not change if we rearrange its terms and write it in the form

$$z_1 z'_1 + z_1 z'_2 + z_2 z'_2 + z_2 z'_1 + z_1 z'_3 + z_2 z'_3 + z_3 z'_3 + z_3 z'_2 + z_3 z'_1 + \cdots \tag{3.21}$$

But the partial sum of the first n^2 terms of (3.21) is easily calculated, and in fact

$$\begin{aligned} z_1 z'_1 &= s_1 s'_1, \\ z_1 z'_1 + z_1 z'_2 + z_2 z'_2 + z_2 z'_1 &= s_2 s'_2, \\ \cdots\cdots\cdots \qquad & \cdots \\ z_1 z'_1 + z_1 z'_2 + \cdots + z_n z'_1 &= s_n s'_n, \end{aligned}$$

where s_n and s'_n denote the partial sums of the first and second of the series (3.18), respectively.[8] It follows that (3.21), and hence (3.20), has the sum

$$\lim_{n \to \infty} s_n s'_n = ss'.$$

Then, because of the associative property (3.17), the series

$$\begin{aligned} z_1 z'_1 + (z_1 z'_2 + z_2 z'_1) + \cdots + (z_1 z'_n + z_2 z'_{n-1} + \cdots + z_n z'_1) + \cdots \\ = \sum_{n=1}^{\infty} (z_1 z'_n + z_2 z'_{n-1} + \cdots + z_n z'_1) \end{aligned} \tag{3.22}$$

---

[7] See e.g., R. C. Buck, *op. cit.*, Theorem 13, p. 120.

[8] The sum of the nine terms written out in (3.21) is $s_3 s'_3$.

has the same sum ss'. It only remains to show that (3.22) is absolutely convergent. But this is an immediate consequence of the inequality

$$|z_1z'_n + z_2z'_{n-1} + \cdots + z_nz'_1| \leqslant |z_1|\,|z'_n| + |z_2|\,|z'_{n-1}| + \cdots + |z_n|\,|z'_1|$$

and the fact that the series

$$|z_1|\,|z'_1| + (|z_1|\,|z'_2| + |z_2|\,|z'_1|) + \cdots + (|z_1|\,|z'_n| + |z_2|\,|z'_{n-1}| + \cdots + |z_n|\,|z'_1|) + \cdots,$$

equal to (3.19), is convergent.

As already noted, an absolutely convergent series has certain commutative and associative properties, i.e., its sum is invariant under rearrangement of terms [cf. (3.12)] and regrouping of terms [cf. (3.17)]. The next result generalizes these properties:

THEOREM 3.8. *Let*

$$\{k_n^{(1)}\}, \{k_n^{(2)}\}, \ldots, \{k_n^{(m)}\}, \ldots$$

be an infinite family of sequences of positive integers such that

$$k_1^{(m)} < k_2^{(m)} < \cdots < k_n^{(m)} < \cdots \qquad (m = 1, 2, \ldots)$$

and such that each positive integer belongs to one and only one of the sequences $\{k_n^{(m)}\}$. *Then, given an absolutely convergent series*

$$\sum_{n=1}^{\infty} z_n \tag{3.23}$$

with sum s, the "iterated series"

$$\sum_{m=1}^{\infty}\left[\sum_{n=1}^{\infty} z_{k_n^{(m)}}\right] \tag{3.24}$$

also converges to s.

Proof. For simplicity, we write

$$z_{k_n^{(m)}} = z_{mn},$$

noting that for every fixed m, the sequence $\{z_{mn}\}$ is a subsequence of $\{z_n\}$. Since (3.23) is absolutely convergent,

$$\sum_{n=1}^{\infty} |z_n| = \sigma,$$

where $\sigma > 0$. It follows that

$$\sum_{n=1}^{q} |z_{mn}| \leqslant \sum_{n=1}^{\infty} |z_n| = \sigma \qquad (q = 1, 2, \ldots).$$

Therefore each of the series

$$\sum_{n=1}^{\infty} z_{mn}$$

is absolutely convergent,[9] and hence convergent, i.e., each of the terms in brackets in (3.24) is well defined. Clearly we have

$$\sum_{m=1}^{p} \left| \sum_{n=1}^{q} z_{mn} \right| \leqslant \sum_{m=1}^{p} \left[\sum_{n=1}^{q} |z_{mn}| \right] \leqslant \sigma \qquad (p, q = 1, 2, \ldots). \quad (3.25)$$

Letting $q \to \infty$ in (3.25), we find that

$$\sum_{m=1}^{p} \left| \sum_{n=1}^{\infty} z_{mn} \right| \leqslant \sigma \qquad (p = 1, 2, \ldots).$$

It follows that the series (3.24) is absolutely convergent, and hence convergent.

To show that the series (3.24) has the same sum as the original series (3.23), we argue as follows: Given any $\varepsilon > 0$, there exists a positive integer $N = N(\varepsilon)$ such that

$$|z_{N+1}| + |z_{N+2}| + \cdots < \varepsilon.$$

Moreover, it is apparent from the structure of (3.24) that there is a positive integer $P = P(N)$ such that the pth partial sum

$$\sum_{m=1}^{p} \left[\sum_{n=1}^{\infty} z_{mn} \right]$$

of the series (3.24) contains all the terms $z_1, \ldots, z_N$ (and an infinite number of other terms besides), provided that $p > P$. Therefore, given any $\varepsilon > 0$, there exists a $P = P(\varepsilon)$ such that

$$\left| s - \sum_{m=1}^{p} \left[\sum_{n=1}^{\infty} z_{mn} \right] \right| \leqslant |z_{N+1}| + |z_{N+2}| + \cdots < \varepsilon$$

provided that $p > P$, i.e.,

$$\lim_{p \to \infty} \sum_{m=1}^{p} \left[\sum_{n=1}^{\infty} z_{mn} \right] = \sum_{m=1}^{\infty} \left[\sum_{n=1}^{\infty} z_{mn} \right] = s,$$

and the theorem is proved.

11. Limit Points of Sets. Bounded Sets

Let E be an arbitrary set of points in the complex plane. A point ζ is said to be a *limit point* of the set E if every neighborhood of ζ contains

---

[9] If $\{c_n\}$ is a bounded sequence of real numbers such that

$$c_1 \leqslant c_2 \leqslant \cdots \leqslant c_n \leqslant \cdots,$$

then $\{c_n\}$ is convergent. See e.g., R. C. Buck, *op. cit.*, Theorem 6, p. 16.

infinitely many (distinct) points of E. Obviously, a set consisting of a finite number of points has no limit points. Suppose the set E consists of the terms of some sequence $\{z_n\}$. Then every point ζ which is a limit point of the set E is also a limit point of the sequence $\{z_n\}$, in the sense of the definition given earlier (p. 28). To see this, we note that if ζ is a limit point of E, any neighborhood of ζ contains infinitely many points of E and hence infinitely many terms of $\{z_n\}$, i.e., a subsequence of $\{z_n\}$. On the other hand, a term ζ' of the sequence $\{z_n\}$ can appear *infinitely often*, i.e., $\{z_n\}$ can have a subsequence $z_{k_1}, z_{k_2}, \ldots, z_{k_n}, \ldots$ all of whose terms equal ζ', without ζ' being a limit point of E (obviously, ζ' will be a limit point of $\{z_n\}$). For example, the points 0 and 1 are both limit points of the sequence

$$0, 1, 0, 1, 0, 1, \ldots, \tag{3.26}$$

but they are not limit points of the set consisting of the elements of (3.26), which contains only the two points 0 and 1. This discrepancy vanishes if we eliminate the possibility of a term appearing infinitely often in a sequence:

THEOREM 3.9. *Let E be an arbitrary set. Then ζ is a limit point of E if and only if there exists a sequence $\{z_n\}$ of distinct points of E converging to ζ.*

Proof. Suppose ζ is a limit point of the set E. We have to find a sequence $\{z_n\}$ of points in E such that $z_j \neq z_k$ if $j \neq k$ and $z_n \to \zeta$ as $n \to \infty$. First consider the neighborhood $\mathcal{N}(\zeta; 1)$ of the point ζ, containing infinitely many points of E, and let z_1 be one of these points. Next consider the neighborhood $\mathcal{N}(\zeta; \frac{1}{2})$, and from the infinitely many points of E in $\mathcal{N}(\zeta; \frac{1}{2})$ choose a point z_2 different from z_1. Suppose repeating this procedure n times leads to n distinct points $z_1, z_2, \ldots, z_n$ of E such that z_k belongs to the neighborhood $\mathcal{N}(\zeta; 1/k)$, $k = 1, 2, \ldots, n$. Then, since the neighborhood $\mathcal{N}(\zeta; 1/(n+1))$ contains infinitely many points of E, one of these points, say z_{n+1}, must be different from the points $z_1, z_2, \ldots, z_n$. In this way, we obtain a sequence $\{z_n\}$ of distinct points of E such that

$$|z_n - \zeta| < \frac{1}{n} \qquad (n = 1, 2, \ldots),$$

i.e., such that $z_n \to \zeta$ as $n \to \infty$.

Conversely, suppose that $\{z_n\}$ is a sequence of distinct points of E such that $z_n \to \zeta$ as $n \to \infty$. Then any neighborhood of the point ζ contains infinitely many points $z_n \in E$, and hence ζ is a limit point of the set E.

Remark 1. Obviously, we can impose the extra requirement that the points of $\{z_n\}$ all be different from the limit point ζ.

Remark 2. The close resemblance between the first part of the proof of Theorem 3.9 and the proof of Theorem 3.3 should be noted.

A set E is said to be *bounded* if there exists a neighborhood of the point $z = 0$ containing all the points of E, i.e., if there exists a $\rho > 0$ such that $|z| < \rho$ for all $z \in E$.

THEOREM 3.10 (*Bolzano-Weierstrass principle for sets*). *Every bounded infinite set E has at least one limit point.*

Proof. Let z_1 be any point of E. Since E is an infinite set, there is a point z_2 of E different from z_1, a point z_3 different from z_1 and z_2, and so on. Suppose we repeat this argument n times, obtaining n distinct points $z_1, z_2, \ldots, z_n$ belonging to E. Then from the infinite set E we can still select another point z_{n+1} different from $z_1, z_2, \ldots, z_n$. Thus there exists a sequence $\{z_n\}$ of distinct points of E which is bounded, like E itself. It follows from Theorem 3.1 that $\{z_n\}$ has at least one limit point ζ, which is also a limit point of E, since any neighborhood of ζ contains infinitely many points $z_n \in E$.

COROLLARY. *Every bounded infinite set E contains a convergent sequence consisting of distinct points of E.*

Proof. In fact, the sequence in question is the subsequence of $\{z_n\}$ converging to ζ, guaranteed by Theorem 3.3.

12. The Limit of a Function of a Complex Variable

THEOREM 3.11. *Let E be an arbitrary infinite set, let $f(z)$ be a function defined on E, and let z_0 be a limit point of E. Suppose the sequence $\{f(z_n)\}$ converges for every sequence of points $\{z_n\}$ such that*[10]

$$z_n \in E, \quad z_n \neq z_0 \qquad (n = 1, 2, \ldots), \qquad z_n \to z_0 \text{ as } n \to \infty. \tag{3.27}$$

Then, if $\{z_n\}$ and $\{z_n'\}$ are two sequences satisfying (3.27),

$$\lim_{n \to \infty} f(z_n) = \lim_{n \to \infty} f(z_n'). \tag{3.28}$$

Proof. The sequence

$$z_1, z_1', z_2, z_2', \ldots, z_n, z_n', \ldots$$

obviously converges to z_0 and consists of points in E different from z_0. Therefore the sequence

$$f(z_1), f(z_1'), f(z_2), f(z_2'), \ldots, f(z_n), f(z_n'), \ldots \tag{3.29}$$

must converge. But this implies (3.28), since any two subsequences of (3.29) must have the same limit.

---

[10] According to Theorem 3.9, at least one such sequence exists.

DEFINITION. *Suppose the sequence* $\{f(z_n)\}$ *converges for every sequence of points* $\{z_n\}$ *satisfying* (3.27), *and let* A *be the common limit (guaranteed by Theorem* 3.11) *of all the sequences* $\{f(z_n)\}$. *Then we say that* $f(z)$ *has the limit* A *at the point* z_0 *(relative to the set* E), *and we write*

$$\lim_{\substack{z \to z_0 \\ z \in E}} f(z) = A. \tag{3.30}$$

THEOREM 3.12. *The function* $f(z)$ *satisfies* (3.30) *if and only if given any* $\varepsilon > 0$, *there exists a number* $\delta = \delta(\varepsilon) > 0$ *such that*

$$|f(z) - A| < \varepsilon \tag{3.31}$$

whenever

$$|z - z_0| < \delta \qquad (z \in E, \quad z \neq z_0). \tag{3.32}$$

Proof. First we prove that the condition is sufficient. Let $\{z_n\}$ be an arbitrary sequence converging to z_0 (where $z_n \in E$, $z_n \neq z_0$), and let $\varepsilon > 0$ be arbitrary. Then, according to (3.31) and (3.32), $|f(z_n) - A| < \varepsilon$ if $n > N[\delta(\varepsilon)] = N'(\varepsilon)$. But this implies (3.30), since $\{z_n\}$ is an arbitrary sequence converging to z_0.

To prove that the condition is necessary, suppose that for some $\varepsilon > 0$ no $\delta > 0$ can be found such that (3.32) implies (3.31). Then, in each neighborhood $\mathcal{N}(z_0; 1/n)$, $n = 1, 2, \ldots$, we can find a point z_n such that

$$|f(z_n) - A| \geqslant \varepsilon > 0. \tag{3.33}$$

Although the sequence $\{z_n\}$ obviously converges to z_0, the inequality (3.33) precludes the possibility of $\{f(z_n)\}$ converging to A. Therefore, (3.30) does not hold, and the proof is complete.

Remark. Our definition of the limit of a function of a complex variable includes as a special case the familiar definition of the limit of a real function of one or two real variables. In fact, if $f(z)$ is real, we can write

$$f(z) = f(x + iy) = u(x, y),$$

where u takes only real values. Let $z_0 = x_0 + iy_0$, and suppose that[11]

$$\lim_{z \to z_0} f(z) = A.$$

Then, according to Theorem 3.12, given any $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$ such that

$$|f(z) - A| = |u(x, y) - A| < \varepsilon \tag{3.34}$$

whenever

$$0 < |z - z_0| = \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta(\varepsilon).$$

[11] Henceforth, for simplicity, we do not explicitly write the extra condition $z \in E$.

But then (3.34) holds whenever

$$0 < |x - x_0| < \delta'(\varepsilon), \qquad 0 < |y - y_0| < \delta'(\varepsilon),$$

where $\delta'(\varepsilon) = \delta(\varepsilon)/\sqrt{2}$, and this is precisely what is meant by writing[12]

$$\lim_{\substack{x \to x_0 \\ y \to y_0}} u(x, y) = A.$$

THEOREM 3.13. *A necessary and sufficient condition for*

$$\lim_{z \to z_0} f(z) = A,$$

where

$$f(z) = u(x, y) + iv(x, y), \qquad A = a + ib,$$

is that

$$\lim_{\substack{x \to x_0 \\ y \to y_0}} u(x, y) = a, \qquad \lim_{\substack{x \to x_0 \\ y \to y_0}} v(x, y) = b.$$

Proof. Use the above remark and Theorem 3.12, together with the same argument as in the proof of Theorem 3.15.

Remark. It follows from Theorem 3.13 that the results concerning algebraic operations on limits of real functions (familiar from elementary calculus) carry over to the case of functions of a complex variable. Thus, if

$$\lim_{z \to z_0} f(z) = A, \qquad \lim_{z \to z_0} g(z) = B,$$

then

$$\lim_{z \to z_0} [f(z) \pm g(z)] = A \pm B,$$

$$\lim_{z \to z_0} f(z)g(z) = AB,$$

$$\lim_{z \to z_0} \frac{f(z)}{g(z)} = \frac{A}{B},$$

where in the last formula it is assumed that $B \neq 0$.

13. Continuous Functions. More Set Theory

Next we introduce the concept of continuity for functions of a complex variable. Let $f(z)$ be a function defined on a set E, and let $z_0 \in E$ be a limit point of E. Then the function $f(z)$ is said to be *continuous at the point* z_0 if

$$\lim_{z \to z_0} f(z) = f(z_0). \tag{3.35}$$

[If $z_0 \in E$ is not a limit point of E, then $f(z)$ is automatically regarded as continuous at z_0.] A function which is continuous at every point of a set E is said to be *continuous on* E. Since, when $f(z)$ takes only real values, our

[12] See e.g., D. V. Widder, *op. cit.*, p. 12.

definition of a limit reduces to the definition familiar from elementary calculus, the same is true of our definition of continuity. In fact, according to Theorem 3.13, if

$$f(z) = u(x, y) + iv(x, y),$$

we can replace (3.35) by the two equivalent relations

$$\lim_{\substack{x \to x_0 \\ y \to y_0}} u(x, y) = u(x_0, y_0), \qquad \lim_{\substack{x \to x_0 \\ y \to y_0}} v(x, y) = v(x_0, y_0),$$

where $z = x + iy$, $z_0 = x_0 + iy_0$. In other words, the complex function $f(z)$ is continuous at the point $z_0 = x_0 + iy_0$ if and only if the real and imaginary parts of $f(z)$, regarded as functions of the two real variables x and y, are continuous at the point (x_0, y_0).

The following properties of continuous functions of a complex variable are immediate consequences of our definition of continuity:

1. If $f(z)$ and $g(z)$ are continuous at the point $z_0 \in E$, then

$$f(z) \pm g(z), \quad f(z)g(z), \quad \frac{f(z)}{g(z)}$$

are also continuous at z_0 [in the last case, it is assumed that $g(z_0) \neq 0$].

2. Suppose the function $f(z)$ with domain E and range $\mathscr{E}$ is continuous at the point $z_0 \in E$, and suppose $\mathscr{E}$ is an infinite set with $w_0 = f(z_0)$ as a limit point. Moreover, suppose the function $\varphi(w)$ has domain $\mathscr{E}$ and is continuous at the point $w_0 \in \mathscr{E}$. Then the *composite function*

$$\varphi[f(z)]$$

is continuous at the point z_0. To see this, we note that if $\{z_n\}$ is any sequence of points in E converging to z_0, then, since $f(z)$ is continuous at z_0, the sequence $\{w_n\}$, where $w_n = f(z_n)$, converges to $w_0 = f(z_0)$. Therefore, since $\varphi(w)$ is continuous at w_0, the sequence $\{\varphi(w_n)\}$ converges to $\varphi(w_0)$. But

$$\varphi(w_n) = \varphi[f(z_n)], \qquad \varphi(w_0) = \varphi[f(z_0)],$$

and hence

$$\lim_{n \to \infty} \varphi[f(z_n)] = \varphi[f(z_0)],$$

i.e.,

$$\lim_{z \to z_0} \varphi[f(z)] = \varphi[f(z_0)],$$

as asserted.

A set F is said to be *closed* if it contains all its limit points.[13] Sets with no limit points (e.g., finite sets, the empty set, the set of all positive integers) are also regarded as closed. The reason for this is apparent from the following

[13] A closed set is often denoted by the letter F (from the French *fermé*).

negative form of the definition of a closed set: A set F is said to be closed if no point which is not a member of F can be a limit point of F. A set F which is both closed and bounded is said to be *compact*.

Example 1. Any finite set is compact.

Example 2. The set of all points in the plane and the set of all positive integers are closed, but not compact.

Example 3. The set F of all points belonging to a given straight line is closed, since any point $z_0 \notin F$ has a neighborhood consisting entirely of points not in F and hence z_0 cannot be a limit point of F. However, F is unbounded and hence not compact.

Example 4. The set F of all points belonging to a given circle is closed, for the same reason as given in Example 3. Moreover, F is obviously bounded and hence compact.

A set O is said to be *open* if each point $z \in O$ has a neighborhood $\mathcal{N}(z)$ consisting entirely of points of O.[14] The empty set is regarded as open.

Example 1. Any neighborhood $\mathcal{N}(z)$ is open.

Example 2. The set of all points in the plane is open.

Example 3. The set of all points which do not belong to a given finite family of lines or circles is open.

THEOREM 3.14 (*Heine-Borel theorem*). *Let F be a compact set, and suppose a neighborhood $\mathcal{N}(z)$ is associated with each point $z \in F$.*[15] *Then there exists a finite number of points $z_1, \ldots, z_n$ in F such that*

$$F \subset \mathcal{N}(z_1) \cup \cdots \cup \mathcal{N}(z_n),$$

i.e., F can be covered by a finite number of neighborhoods $\mathcal{N}(z)$.

Proof. The proof closely resembles that of Theorem 3.1. Let Q_0 be a square with sides parallel to the coordinate axes and with center at the origin, which contains the set F (cf. Figure 3.1). Suppose the theorem is false, i.e., suppose F cannot be covered by a finite number of neighborhoods $\mathcal{N}(z)$. The coordinate axes divide Q_0 into four squares, and correspondingly, the set F is divided into four subsets, i.e., the intersections of F with each of these squares. At least one of the four subsets of F cannot be covered by a finite number of neighborhoods $\mathcal{N}(z)$. Let Q_1 be the square containing this subset. Similarly, dividing Q_1 into four congruent subsquares, we can find a new square, say Q_2,

[14] An open set is often denoted by the letter O (from the French *ouvert*).

[15] The set of neighborhoods $\mathcal{N}(z)$, $z \in F$ may be uncountable (see Problem 3.2).

whose intersection with F cannot be covered by a finite number of neighborhoods $\mathcal{N}(z)$. Continuing this construction, we arrive at a sequence of squares

$$Q_0 \supset Q_1 \supset Q_2 \supset \cdots \supset Q_n \supset \cdots,$$

each strictly including the next, such that $Q_n \cap F$ cannot be covered by a finite number of neighborhoods $\mathcal{N}(z)$. As in the proof of Theorem 3.1, let ζ be the unique point belonging to every square Q_n. Every Q_n contains an infinite number of points of F, for otherwise we could immediately find a trivial finite covering of $Q_n \cap F$. Therefore ζ must be a limit point of F, since any neighborhood of ζ contains all the squares Q_n starting from some value of n, and hence contains infinitely many points of F. Since F is closed, ζ must belong to F, and hence there is a neighborhood containing ζ in the original set of neighborhoods, i.e., $\mathcal{N}(\zeta)$. But as just noted, starting from some value of n we have $Q_n \subset \mathcal{N}(\zeta)$ and hence $(Q_n \cap F) \subset \mathcal{N}(\zeta)$, i.e., $Q_n \cap F$ is covered by the single neighborhood $\mathcal{N}(\zeta)$, contrary to hypothesis. This contradiction proves the theorem.

Remark. Obviously, the theorem remains true if $\mathcal{N}(z)$ is replaced by any neighborhood containing z (not necessarily one with z as its center), or more generally by any open set containing z (cf. footnote 1, p. 27).

THEOREM 3.15. *Let F be a compact set, and let $f(z)$ be a continuous function on F. Then the set $f(F)$ is compact, i.e., the continuous image of a compact set is compact.*

Proof. First we prove that $\mathcal{F} = f(F)$ is bounded. Since $f(z)$ is continuous on F, given any $\varepsilon > 0$, we can associate a neighborhood $\mathcal{N}(z)$ with each $z \in F$ such that

$$|f(z) - f(z')| < \varepsilon$$

whenever $z' \in \mathcal{N}(z) \cap F$. In other words,

$$f(z') \in \mathcal{N}(f(z); \varepsilon)$$

whenever $z' \in \mathcal{N}(z) \cap F$. According to Theorem 3.14, there exists a finite number of points $z_1, \ldots, z_n$ in F such that

$$F \subset \mathcal{N}(z_1) \cup \cdots \cup \mathcal{N}(z_n),$$

and hence

$$\mathcal{F} \subset \mathcal{N}(f(z_1); \varepsilon) \cup \cdots \cup \mathcal{N}(f(z_n); \varepsilon),$$

i.e., $\mathcal{F}$ can be covered by a finite number of neighborhoods and hence is bounded.

Next we prove that $\mathcal{F}$ is closed. Let ω be a limit point of $\mathcal{F}$. Then, according to Theorem 3.9, there is a sequence $\{w_n\}$ of distinct points of $\mathcal{F}$ converging to ω. For each n, let z_n be a point in F such that $f(z_n) = w_n$,

and consider the sequence $\{z_n\}$. Since the sequence $\{w_n\}$ consists of distinct points, so does $\{z_n\}$. Since $\{z_n\}$ is bounded (being a subset of the compact set F), $\{z_n\}$ has at least one limit point ζ, according to Theorem 3.1 (or Theorem 3.10), and since F is closed, ζ must belong to F. By the corollary to Theorem 3.10, $\{z_n\}$ contains a subsequence $\{z_{k_n}\}$ converging to $\zeta \in F$. Moreover, since $f(z)$ is continuous on F,

$$\lim_{n \to \infty} f(z_{k_n}) = f(\zeta). \tag{3.36}$$

On the other hand, $\{f(z_{k_n})\}$ is a subsequence of $\{w_n\}$, and hence itself converges to ω:

$$\lim_{n \to \infty} f(z_{k_n}) = \omega. \tag{3.37}$$

It follows by comparing (3.36) and (3.37) that $\omega = f(\zeta)$, i.e., that $\omega \in \mathscr{F}$ and $\mathscr{F}$ is closed, as asserted.

COROLLARY. *Let F be a compact set, and let $f(z)$ be a continuous function on F. Then there exists a constant $M > 0$ such that*

$$|f(z)| \leqslant M$$

for all $z \in F$, i.e., $f(z)$ is bounded on F.[16]

Proof. If $f(z)$ is continuous on F, so is $|f(z)|$, since

$$\big||f(z)| - |f(z_0)|\big| \leqslant |f(z) - f(z_0)|$$

[cf. (2.6)].

THEOREM 3.16. *Let F be a compact set, and let $f(z)$ be a continuous function on F. Then there exist points z_0, Z_0 in F such that*

$$|f(z_0)| \leqslant |f(z)| \leqslant |f(Z_0)|$$

for all $z \in F$, i.e., $|f(z)|$ achieves its greatest lower bound and its least upper bound on F.[17]

Proof. Since $|f(z)|$ is continuous on F, the set $\mathscr{F}$ of real numbers

$$w = |f(z)|, \quad z \in F$$

is compact, according to Theorem 3.15. Let m and M be the greatest lower bound and the least upper bound, respectively, of $\mathscr{F}$ (these exist

---

[16] Equivalently, a function $f(z)$ is said to be *bounded* on a set E, if $f(E)$ is a bounded set (cf. p. 44).

[17] An equivalent version of Theorem 3.16 is the following: *Let F be a compact set, and let $f(z)$ be a continuous function on F. Then there exist points z_0, Z_0 in F such that*

$$|f(z_0)| = \min_{z \in F} |f(z)|, \qquad |f(Z_0)| = \max_{z \in F} |f(z)| .$$

since $\mathscr{F}$ is bounded). By the very definition of m, given any $\varepsilon > 0$, the neighborhood $\mathscr{N}(m; \varepsilon)$ contains at least one point of $\mathscr{F}$ (which may be m itself). Therefore, m is either already known to be in $\mathscr{F}$, or else m is a limit point of $\mathscr{F}$, and hence an element of $\mathscr{F}$ (since $\mathscr{F}$ is closed). By a similar argument, $M \in \mathscr{F}$. In other words, there exist points z_0, Z_0 in F such that

$$f(z_0) = m, \qquad f(Z_0) = M.$$

THEOREM 3.17. *Let F be a compact set, and let $f(z)$ be a continuous function on F. Then, given any $\varepsilon > 0$, there exists a number $\delta = \delta(\varepsilon) > 0$ such that*

$$|f(z) - f(z')| < \varepsilon$$

for every pair of points z, z' in F such that

$$|z - z'| < \delta, \tag{3.38}$$

i.e., $f(z)$ is uniformly continuous on F.

Proof. The important point here is that one choice of $\delta(\varepsilon)$ suffices for *all* z, $z' \in F$. Since $f(z)$ is continuous on F, given any $\varepsilon > 0$, we can associate a neighborhood $\mathscr{N}(z; \delta(z))$ with each $z \in F$ such that

$$|f(z) - f(z')| < \tfrac{1}{2}\varepsilon$$

whenever $z' \in \mathscr{N}(z; \delta(z))$. Now consider the set of "half-size" neighborhoods $\mathscr{N}(z; \frac{1}{2}\delta(z))$, $z \in F$. According to Theorem 3.14, there exists a finite number of points $z_1, \ldots, z_n$ in F such that

$$F \subset \mathscr{N}(z_1; \tfrac{1}{2}\delta_1) \cup \cdots \cup \mathscr{N}(z_n; \tfrac{1}{2}\delta_n), \tag{3.39}$$

where $\delta_1 = \delta(z_1), \ldots, \delta_n = \delta(z_n)$. Then a suitable δ for use in the inequality (3.38) is given by

$$\delta = \tfrac{1}{2} \min (\delta_1, \ldots, \delta_n).$$

To see this, let z and z' be any two points in F such that $|z - z'| < \delta$. According to (3.39), z belongs to some neighborhood $\mathscr{N}(z_k; \frac{1}{2}\delta_k)$, $1 \leqslant k \leqslant n$, and hence

$$|f(z) - f(z_k)| < \tfrac{1}{2}\varepsilon,$$

since obviously $\mathscr{N}(z_k; \frac{1}{2}\delta_k) \subset \mathscr{N}(z_k; \delta_k)$. Moreover, z' belongs to $\mathscr{N}(z_k; \delta_k)$, since

$$|z' - z_k| = |(z' - z) + (z - z_k)| \leqslant |z' - z| + |z - z_k| < \delta + \tfrac{1}{2}\delta_k \leqslant \delta_k.$$

It follows that

$$|f(z') - f(z_k)| < \tfrac{1}{2}\varepsilon,$$

and hence

$$\begin{aligned} |f(z) - f(z')| &= |[f(z) - f(z_k)] + [f(z_k) - f(z')]| \\ &\leqslant |f(z) - f(z_k)| + |f(z') - f(z_k)| < \varepsilon, \end{aligned}$$

as required.

14. The Distance between Two Sets

Let E be a nonempty set, and let z be any point in the complex plane. By the *distance between z and E* we mean the quantity

$$\rho(z, E) = \inf_{z' \in E} |z - z'|, \tag{3.40}$$

i.e., the greatest lower bound of the distances between z and all the points of E. If $\rho(z, E) > 0$, then obviously z cannot belong to E, but if $\rho(z, E) = 0$, z may or may not belong to E. In the latter case, E contains points arbitrarily close to z, i.e., z is a limit point of E. Since a closed set contains all its limit points, we have proved

THEOREM 3.18. *If E is a closed set and z is a point not in E, then* $\rho(z, E) > 0$.

Now let D and E be two nonempty sets. By the *distance between D and E* we mean the quantity

$$\rho(D, E) = \inf_{\substack{z \in D \\ z' \in E}} |z - z'|, \tag{3.41}$$

i.e., the greatest lower bound of the distances between all pairs of points z and z', with $z \in D$, $z' \in E$. If D consists of a single point z, then (3.41) reduces to (3.40). Moreover, it is obvious that $\rho(D, E) = \rho(E, D)$. If $\rho(D, E) > 0$, then D and E must be disjoint, but if $\rho(D, E) = 0$, D and E may or may not have a point in common. In the latter case, there are pairs of points z and z', with $z \in D$, $z' \in E$, such that $|z - z'|$ is arbitrarily small. Clearly, this is possible when one of the sets is not closed,[18] but it is even possible when D and E are both closed. For example, let D consist of the points belonging to a hyperbola, and let E consist of the points belonging to the asymptotes of the hyperbola. In this case, D and E are both closed and $D \cap E = 0$, but $\rho(D, E) = 0$. Actually, the appropriate generalization of Theorem 3.18 is given by

[18] If D is not closed, let E consist of the limit points of D which do not belong to D. Then $D \cap E = 0$ and $\rho(D, E) = 0$.

THEOREM 3.19. *If D and E are two disjoint closed sets, and if one of the sets, say D, is bounded, then $\rho(D, E) > 0$.*

Proof. If $z \in D$, then $z \notin E$ and hence $\rho(z, E) > 0$, by Theorem 3.18. With each point $z \in D$ we associate the "half-size" neighborhood $\mathcal{N}(z; \frac{1}{2}\rho(z, E))$. According to Theorem 3.14, since D is compact, there exists a finite number of points $z_1, \ldots, z_n$ in D such that

$$D \subset \mathcal{N}(z_1; \tfrac{1}{2}\rho_1) \cup \cdots \cup \mathcal{N}(z_n; \tfrac{1}{2}\rho_n), \tag{3.42}$$

where $\rho_1 = \rho(z_1, E), \ldots, \rho_n = \rho(z_n, E)$. Let

$$\delta = \tfrac{1}{2} \min (\rho_1, \ldots, \rho_n),$$

and consider any point $z \in D$. According to (3.42), z belongs to some neighborhood $\mathcal{N}(z_k; \frac{1}{2}\rho_k)$, $1 \leqslant k \leqslant n$, and since $\mathcal{N}(z_k; \rho_k)$ contains no points of E, if z' is any point of E, we have

$$|z - z'| \geqslant \tfrac{1}{2}\rho_k \geqslant \delta.$$

Therefore,

$$\rho(D, E) = \inf_{\substack{z \in D \\ z' \in E}} |z - z'| \geqslant \delta > 0,$$

as asserted.

THEOREM 3.20. *If E is a compact set and z is an arbitrary point, then there exists a point ζ in E such that*

$$\rho(z, E) = |z - \zeta|.$$

Proof. If $z \in E$, $\rho(z, E) = 0$ and the theorem is trivial. Therefore suppose $z \notin E$. The function $f(z') = |z - z'|$, $z' \in E$ is continuous on E, and hence, by Theorem 3.16, $f(z')$ achieves its greatest lower bound on E at some point $\zeta \in E$, i.e.,

$$|z - z'| \geqslant |z - \zeta| \tag{3.43}$$

for all $z' \in E$. It follows from (3.43) that

$$\rho(z, E) = \inf_{z' \in E} |z - z'| = |z - \zeta|,$$

as required.

THEOREM 3.21. *If D and E are two disjoint compact sets, then there exists a pair of points z_1 and z_2, with $z_1 \in D$, $z_2 \in E$, such that*

$$\rho(D, E) = |z_1 - z_2|.$$

Proof. First we prove that $\rho(z, E)$ is a continuous function of z. According to Theorem 3.20, there exists a point $\zeta \in E$ such that

$$\rho(z, E) = |z - \zeta|,$$

where, in general, ζ depends on z. Then, given any point z', we have

$$|z' - \zeta| = |(z' - z) + (z - \zeta)| \leqslant |z' - z| + |z - \zeta| = |z' - z| + \rho(z, E).$$

Therefore

$$\rho(z', E) \leqslant |z' - \zeta| \leqslant |z' - z| + \rho(z, E),$$

or

$$\rho(z', E) - \rho(z, E) \leqslant |z' - z|. \tag{3.44}$$

Interchanging the roles of z and z' in (3.46), we obtain

$$\rho(z, E) - \rho(z', E) \leqslant |z - z'|. \tag{3.45}$$

Together, (3.44) and (3.45) imply

$$|\rho(z, E) - \rho(z', E)| \leqslant |z - z'|,$$

and hence to make $|\rho(z, E) - \rho(z', E)| < \varepsilon$, we need only choose $|z - z'| < \delta(\varepsilon) = \varepsilon$, i.e., $\rho(z, E)$ is a continuous function of z, as asserted.

In particular, the function $f(z) = \rho(z, E)$, $z \in D$ is continuous on D, and hence, by Theorem 3.16, $f(z)$ achieves its greatest lower bound on D at some point $z_1 \in D$, i.e.,

$$\rho(z, E) \geqslant \rho(z_1, E)$$

for all $z \in D$. But then

$$|z - z'| \geqslant \rho(z, E) \geqslant \rho(z_1, E) = |z_1 - z_2| \tag{3.46}$$

for any pair of points $z \in D$, $z' \in E$, where z_2 is the point in E for which the equality in (3.46) holds (see Theorem 3.20). It follows from (3.46) that

$$\rho(D, E) = \inf_{\substack{z \in D \\ z' \in E}} |z - z'| = |z_1 - z_2|,$$

as required.

PROBLEMS

3.1. Let Σ be a set whose members are themselves sets. Prove the relations

$$D \cap \Big(\bigcup_{E \in \Sigma} E\Big) = \bigcup_{E \in \Sigma} (D \cap E), \qquad D \cup \Big(\bigcap_{E \in \Sigma} E\Big) = \bigcap_{E \in \Sigma} (D \cup E),$$

where D is an arbitrary set. Prove the *De Morgan formulas*

$$D - \bigcup_{E \in \Sigma} E = \bigcap_{E \in \Sigma} (D - E), \qquad D - \bigcap_{E \in \Sigma} E = \bigcup_{E \in \Sigma} (D - E),$$

and specialize these formulas to the case where D is the whole complex plane.

3.2. By a *finite* set is meant a set consisting of a finite number of elements, and by an *infinite* set is meant a set which is not finite. (The empty set is regarded as finite.) Two sets are said to be *equivalent* if a one-to-one correspondence (see p. 26) can be established between them. A set is said to be

countably infinite if it is equivalent to the set of all positive integers. By a *countable* set is meant either a finite set or a countably infinite set. A set which is not countable is said to be *uncountable*.

Prove that

a) The set of all rational numbers is countable;
b) The set of all real numbers is uncountable;
c) Any subset of a countable set is countable;
d) The union of countably many countable sets is countable.

3.3. Prove that if $z_n \to \zeta$ as $n \to \infty$, then

$$\lim_{n \to \infty} \frac{z_1 + \cdots + z_n}{n} = \zeta.$$

More generally, prove that if $z_n \to \zeta$ as $n \to \infty$, then

$$\lim_{n \to \infty} \frac{c_1 z_1 + \cdots + c_n z_n}{c_1 + \cdots + c_n} = \zeta,$$

where $c_1, c_2, \ldots, c_n, \ldots$ is any sequence of positive numbers such that

$$\lim_{n \to \infty} (c_1 + \cdots + c_n) = +\infty.$$

3.4. Find the limit, if any, of each of the following sequences:

a) $\left\{\frac{2^n}{n!} + \frac{in}{2^n}\right\}$; b) $\{\sqrt[n]{n} + inq^n\} (|q| < 1)$; c) $\left\{\sqrt[n]{a} + i \sin \frac{1}{n}\right\} (a > 0)$.

3.5. Prove that if the sequence $\{z_n\}$ converges to a nonzero limit ζ which is not a negative real number, then $\arg z_n \to \arg \zeta$, where, as on p. 35, we ignore the finite number of points where z_n may vanish.

3.6. Suppose the sequence $\{z_n\}$ converges to a nonzero limit ζ, and let $\{\Phi_n\}$ be any sequence of values of $\operatorname{Arg} z_n$ satisfying the inequality

$$|\Phi_m - \Phi_n| < \pi$$

if both m and n exceed some positive integer N. Prove that $\{\Phi_n\}$ converges to one of the values of $\operatorname{Arg} \zeta$.

3.7. Give an example of a series which can be shown to converge by using the root test, but not by using the ratio test (see p. 38).

3.8. Let $\{a_n\}$ be a sequence whose partial sums are bounded, and let $\{b_n\}$ be a sequence such that $b_n \to 0$ as $n \to \infty$ and

$$b_1 \geqslant b_2 \geqslant \cdots \geqslant b_n \geqslant \cdots.$$

Prove that

$$\sum_{n=1}^{\infty} a_n b_n$$

converges.

Hint. Use Theorem 3.6.

3.9. Let the sequence $\{b_n\}$ be the same as in Problem 3.8. Prove that

$$\sum_{n=1}^{\infty} (-1)^{n+1} b_n$$

converges.

3.10. According to Problem 3.9, the series

$$\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{\sqrt{n}}$$

converges. Prove that this series does not converge absolutely. Moreover, show that the product of the series with itself, i.e., the series

$$\sum_{n=1}^{\infty} (-1)^{n+1} (b_1 b_n + b_2 b_{n-1} + \cdots + b_n b_1)$$

is divergent (see Theorem 3.7).

Hint. $\dfrac{1}{\sqrt{1}\sqrt{n}} + \dfrac{1}{\sqrt{2}\sqrt{n-1}} + \cdots + \dfrac{1}{\sqrt{n}\sqrt{1}} > \dfrac{n}{\sqrt{n}\sqrt{n}} = 1.$

3.11. Investigate the convergence of the following series:

a) $\sum_{n=0}^{\infty} \dfrac{1}{(1+i)^n}$; b) $\sum_{n=0}^{\infty} \left(\cos \dfrac{n\pi}{4} + i \sin \dfrac{n\pi}{4}\right)$;

c) $\sum_{n=0}^{\infty} \dfrac{1}{(1+i)^n} \left(\cos \dfrac{n\pi}{4} + i \sin \dfrac{n\pi}{4}\right)$; d) $\sum_{n=0}^{\infty} \dfrac{(1+i)^n}{i(1+i)^n + 2}$.

3.12. Prove that if the series

$$\sum_{n=1}^{\infty} z_n$$

converges absolutely, then so does the series

$$\sum_{n=1}^{\infty} z_n^2.$$

Is the converse true?

3.13. Prove that if the series

$$\sum_{n=1}^{\infty} z_n$$

converges, and if

$$|\arg z_n| \leqslant \alpha < \frac{\pi}{2}$$

for all n, then the series converges absolutely.

3.14. Give an example of a series

$$\sum_{n=1}^{\infty} z_n$$

for which the "iterated series" (3.24) depends on the choice of the family of sequences $\{k_n^{(m)}\}$.

3.15. Find the limit points of the set of all points z such that

a) $z = 1 + (-1)^n \dfrac{n}{n+1}$ $(n = 1, 2, \ldots)$;

b) $z = \dfrac{1}{m} + \dfrac{i}{n}$ $(m, n = \pm 1, \pm 2, \ldots)$;

c) $z = \dfrac{p}{m} + i\dfrac{q}{n}$ $(m, n, p, q = \pm 1, \pm 2, \ldots)$; d) $|z| < 1$.

3.16. Given an arbitrary set E, a point $z \in E$ is said to be *isolated* if z is not a limit point of E, i.e., if there is a neighborhood of z containing no point of E other than z itself. Prove that the set of all isolated points of E is countable (see Problem 3.2).

3.17. Give an example of a function $f(z)$ which has a limit A at a point z_0 relative to every straight line passing through z_0, but not relative to a neighborhood of z_0.

3.18. The functions

$$\frac{\operatorname{Re} z}{|z|}, \quad \frac{z}{|z|}, \quad \frac{\operatorname{Re} z^2}{|z|^2}, \quad \frac{z \operatorname{Re} z}{|z|}$$

are all defined for $z \neq 0$. Which of them can be defined at the point $z = 0$ in such a way that the extended functions are continuous at $z = 0$?

Ans. Only $f(z) = \dfrac{z \operatorname{Re} z}{|z|}$, $f(0) = 0$.

3.19. Prove that

a) A set is closed if and only if its complement is open;
b) A set is open if and only if its complement is closed.

3.20. Prove that

a) The intersection of an arbitrary number of closed sets is closed;
b) The union of a finite number of closed sets is closed.

3.21. Prove that

a) The union of an arbitrary number of open sets is open;
b) The intersection of a finite number of open sets is open.

3.22. Give examples showing why the word "finite" is needed in Problems 3.20(b) and 3.21(b).

3.23. A set E is said to have the *Heine-Borel property* (cf. Theorem 3.14) if given any set of neighborhoods $\mathcal{N}(z)$, $z \in E$, there exists a finite number of points $z_1, \ldots, z_n$ in E such that

$$E \subset \mathcal{N}(z_1) \cup \cdots \cup \mathcal{N}(z_n).$$

Prove the equivalence of the following three statements:

a) E has the Heine-Borel property;
b) Every infinite subset of E has a limit point in E;
c) E is compact, i.e., closed and bounded.

Hint. In Theorem 3.14 we have already proved that c) implies a), and the fact that c) implies b) is an immediate consequence of Theorem 3.10. Two further implications must be proved.

Comment. Thus, we might just as well have defined a compact set as a set with the Heine-Borel property, or as a set with property b).

3.24. Give an example of a closed set F whose continuous image $f(F)$ is not closed.

Hint. In view of Theorem 3.15, F must be unbounded.

3.25. Let K be the open unit disk $|z| < 1$. Are the functions

$$\frac{1}{1-z}, \qquad \frac{1}{1+z^2}$$

continuous on K, and if so, are they uniformly continuous on K?

3.26. Suppose the function $f(z)$ is uniformly continuous on the open unit disk $|z| < 1$. Let C be the unit circle $|z| = 1$, let ζ be a point of C, and let $\{z_n\}$ be a sequence of points in K such that $z_n \to \zeta$ as $n \to \infty$. Prove that

a) $\lim_{n \to \infty} f(z_n)$ exists and is independent of the choice of $\{z_n\}$;

b) If $f(z)$ is defined on C by making the indicated passage to the limit at every point of C, then $f(z)$ is continuous on the closed unit disk $|z| \leqslant 1$.

3.27. Let E be a bounded set. Then by the *diameter* of E we mean the quantity

$$\operatorname{diam}(E) = \sup_{z, z' \in E} |z - z'|,$$

i.e., the least upper bound of the distances between pairs of points in E. Prove that if E is compact, there exists a pair of points z_1 and z_2 in E such that

$$\operatorname{diam}(E) = |z_1 - z_2|.$$

3.28. Prove that in Theorems 3.20 and 3.21 we can relax the condition that E be compact, by requiring only that E be closed.

4

CONNECTEDNESS. CURVES AND DOMAINS

15. Connected Sets. Continuous Curves and Continua

Given a set E, by the *derived set* of E, denoted by E^d, we mean the set of all limit points of E. Obviously, E is closed if and only if $E^d \subset E$. A set E is said to be *connected* if given any decomposition of E into two nonempty disjoint sets E_1 and E_2, at least one of the sets E_1 and E_2 contains a limit point of the other. In other words, E is connected if and only if

$$E_1 \cup E_2 = E, \quad E_1 \cap E_2 = 0, \quad E_1 \neq 0, \quad E_2 \neq 0 \tag{4.1}$$

implies that $E_1^d \cap E_2 \neq 0$ or $E_1 \cap E_2^d \neq 0$. Equivalently, E is connected if and only if there do not exist sets E_1 and E_2 such that (4.1) holds and $E_1^d \cap E_2 = E_1 \cap E_2^d = 0$. Clearly, the empty set and any set consisting of a single point are connected.

Example 1. Any set consisting of a finite number of points $z_1, \ldots, z_n$ ($n > 1$) is *disconnected* (i.e., not connected).

Example 2. More generally, the union of a finite number of disjoint closed sets $F_1, \ldots, F_n$ ($n > 1$) is disconnected.

We now study a very important class of compact connected sets, i.e., the sets of points belonging to continuous curves (see Theorem 4.3 below). A complex function $z = f(t)$ of a real variable which is defined and continuous on a closed interval $a \leqslant t \leqslant b$ is said to define a (*continuous*) *curve*. The

values of the function are called the *points* of the curve, and the set of points in the curve, i.e., the range of $f(t)$, is often referred to simply as the *curve* (when no ambiguity can arise). In particular, the points $z_0 = f(a)$ and $z_1 = f(b)$ are called the *initial* and *final* points of the curve, respectively, and z_0, z_1 are called the *end points* of the curve. The initial and final points of a curve may coincide, in which case the curve is said to be *closed* (not to be confused with the concept of a closed set). The real variable t is called the *parameter* of the curve, and the equation $z = f(t)$, mapping the values of the parameter onto the points of the curve, is called the (*parametric*) *equation* of the curve.

Remark. In calling $z_0 = f(a)$ the *initial* point and $z_1 = f(b)$ the *final* point of the curve L with equation $z = f(t)$, $a \leqslant t \leqslant b$, we ascribe an *orientation* to L, whereby a point $z' = f(t') \in L$ is regarded as *preceding* a point $z'' = f(t'') \in L$ if $z' \neq z''$ and $t' < t''$. We shall always regard L as being traversed in the direction of increasing t, a fact some authors emphasize by calling L an *oriented curve*.

Two curves L and Λ, with equations $z = f(t)$, $a \leqslant t \leqslant b$ and $z = \varphi(\tau)$, $\alpha \leqslant \tau \leqslant \beta$, respectively, are regarded as identical (except possibly for direction) if the equation of one curve can be transformed into the equation of the other by means of a *continuous strictly monotonic* change of parameter, i.e., if there exists a continuous strictly monotonic function $\tau = \lambda(t)$, $a \leqslant t \leqslant b$ (see Problem 4.7), such that the function $\varphi[\lambda(t)]$ coincides with $f(t)$ on the interval $[a, b]$.[1] Then we say that the two curves have the *same* direction if the function $\lambda(t)$ is *increasing*, or *opposite* directions if $\lambda(t)$ is *decreasing*. In the latter case, the initial point of L is the same as the final point of Λ, and *vice versa*. The curve differing from L only by the direction in which it is traversed will be denoted by $-L$.

If the same point z corresponds to more than one parameter value in the half-open interval $a \leqslant t < b$, we say that z is a *multiple point* of the curve $z = f(t)$, $a \leqslant t \leqslant b$. A curve with no multiple points is called a *Jordan curve*. It is to allow for the possibility of *closed* Jordan curves, i.e., Jordan curves whose end points coincide, that we consider parameter values in the *half-open* interval $a \leqslant t < b$.[2]

Example 1. The functions

$$z = t, \quad z = t^2, \quad z = 1 - t \qquad (0 \leqslant t \leqslant 1) \tag{4.2}$$

---

[1] Then, of course, $f[\lambda^{-1}(\tau)]$ also coincides with $\varphi(\tau)$ on the interval $[\alpha, \beta]$, where $\lambda^{-1}(\tau)$ is the (continuous) inverse of the function $\lambda(t)$ [see Problems 4.8 and 6.7].

[2] In the literature, a curve $z = f(t)$, $a \leqslant t \leqslant b$, is often called an *arc* if $f(a) \neq f(b)$, and a *simple arc* or a *Jordan arc* if in addition $f(t)$ is one-to-one on $a \leqslant t \leqslant b$. Moreover, if $f(a) = f(b)$ and $f(t)$ is one-to-one on $a \leqslant t < b$, the curve is often called a *simple closed curve* or just a *Jordan curve* (without the extra adjective *closed*).

all define the same continuous curve (except for direction), corresponding to the segment of the real axis lying between the points $z = 0$ and $z = 1$. The first and second representations of the curve have the same direction (with initial point $z = 0$ and final point $z = 1$), but the third representation has the opposite direction (with initial point $z = 1$ and final point $z = 0$). Obviously, the curve (4.2) has no multiple points and hence is a Jordan curve. Since its initial and final points do not coincide, (4.2) is not a closed Jordan curve.

Example 2. The functions

$$z = t, \quad z = \sin \pi t \qquad (0 \leqslant t \leqslant 1) \tag{4.3}$$

define different continuous curves L and Λ, since there is no continuous strictly monotonic change of parameter carrying one of these curves into the other. In fact, such a transformation would have to carry the monotonic function $f(t) = t$ into another monotonic function, but the function $\varphi(t) = \sin \pi t$ is not monotonic in the interval [0, 1]. We note that although the two curves (4.3) are different, they consist of the same set of points, i.e., the points in the interval [0, 1]. Therefore, the fact that two continuous curves consist of the same points is not a sufficient condition (but obviously a necessary condition) for the curves to be identical. We also note the following facts:

1. The curve $z = t$ is a Jordan curve, but not a closed curve;
2. The curve $z = \sin \pi t$ is not a Jordan curve, since it has multiple points
$$\sin \pi t = \sin \pi(1 - t) \qquad (0 < t < \tfrac{1}{2});$$
3. The curve $z = \sin \pi t$ is a closed curve, since its initial and final points coincide, i.e., $\sin \pi \cdot 0 = \sin \pi \cdot 1 = 0$.

Example 3. Let $\Delta_1, \ldots, \Delta_n$ be a family of (straight) line segments in the plane, with definite directions, such that the final point of each segment Δ_k $(k < n)$ coincides with the initial point of the next segment Δ_{k+1}. Let $a_1, \ldots, a_n$ be the complex numbers represented by the vectors $\Delta_1, \ldots, \Delta_n$, respectively, and let z_0 be the initial point of the segment Δ_1. Then with the given family of segments we can associate the following continuous curve, defined on the interval $0 \leqslant t \leqslant n$:

$$z = z_0 + a_1 + \cdots + a_{k-1} + a_k(t - k + 1) \qquad (k - 1 \leqslant t \leqslant k; k = 1, \ldots, n).$$

A curve of this type is called a *polygonal curve*, with segments $\Delta_1, \ldots, \Delta_n$. Whether or not a polygonal curve is closed depends on whether or not the final point of Δ_n coincides with the initial point of Δ_1. A polygonal curve is a Jordan curve if and only if it does not intersect itself, which means that its segments have no points in common except those permitted by the construction, i.e., the final point of Δ_k $(k < n)$ coincides with the initial point of

Δ_{k+1}, and the final point of Δ_n may coincide with the initial point of Δ_1 (if the curve is closed).

Example 4. The circle

$$z = z_0 + r(\cos t + i \sin t) \qquad (0 \leqslant t \leqslant 2\pi),$$

with center z_0 and radius r, is a closed Jordan curve.

By a *continuum* we mean a nonempty bounded set which is closed and connected (i.e., compact and connected). With this definition, a set consisting of a single point is a continuum. A set E is said to be *arcwise connected* if every pair of points $z, z' \in E$ can be joined by a continuous curve consisting only of points of E, with one of the points z, z' as its initial point and the other as its final point.

THEOREM 4.1. *A nonempty compact set F is connected (and hence is a continuum) if given any two points z, z' in F and any $\varepsilon > 0$, there exists a finite number of points $z = z_0, z_1, \ldots, z_n = z'$ in F such that*

$$|z_{k-1} - z_k| < \varepsilon \qquad (k = 1, 2, \ldots, n). \tag{4.4}$$

Proof. Suppose (4.4) holds but F is disconnected. Then there are two nonempty disjoint sets F_1 and F_2 such that

$$F_1 \cup F_2 = F, \qquad F_1^d \cap F_2 = F_1 \cap F_2^d = 0. \tag{4.5}$$

Since F is closed,

$$F_1^d \subset F, \qquad F_2^d \subset F,$$

and therefore by (4.5)

$$F_1^d \subset F_1, \qquad F_2^d \subset F_2,$$

i.e., F_1 and F_2 are closed and hence compact. It follows by Theorem 3.19 that $\rho(F_1, F_2) > 0$. Therefore, choosing $z \in F_1$ and $z' \in F_2$, we cannot find a finite number of points $z = z_0, z_1, \ldots, z_n = z'$ in F such that

$$|z_{k-1} - z_k| < \rho(F_1, F_2) \qquad (k = 1, 2, \ldots, n).$$

In fact, among these points there must be two with consecutive indices, say $z_{\kappa-1}$ and z_κ, such that the first belongs to F_1 and the second to F_2, which implies that

$$|z_{\kappa-1} - z_\kappa| \geqslant \rho(F_1, F_2).$$

This contradiction proves the theorem.

COROLLARY. *Every line segment is connected.*

THEOREM 4.2. *Let E be a connected set and let $f(z)$ be a continuous function on E. Then the set $f(E)$ is connected, i.e., the continuous image of a connected set is connected.*

Proof. Suppose $\mathscr{E} = f(E)$ is disconnected, i.e., suppose there exists two nonempty disjoint sets $\mathscr{E}_1$ and $\mathscr{E}_2$ such that

$$\mathscr{E}_1 \cup \mathscr{E}_2 = \mathscr{E}, \qquad \mathscr{E}_1^d \cap \mathscr{E}_2 = \mathscr{E}_1 \cap \mathscr{E}_2^d = 0.$$

Let E_1 consist of all points z in E such that $f(z) \in \mathscr{E}_1$, and let E_2 consist of all points z in E_2 such that $f(z) \in \mathscr{E}_2$. Then E_1 and E_2 are disjoint and nonempty, and $E_1 \cup E_2 = E$. Moreover, neither of the sets E_1, E_2 contains a limit point of the other, i.e., $E_1^d \cap E_2 = E_1 \cap E_2^d = 0$. In fact, if $\zeta \in E_1$ were a limit point of E_2, then by Theorem 3.9, E_2 would contain a sequence $\{z_n\}$ of (distinct) points converging to ζ. But since $f(z)$ is continuous on E, it would follow that

$$\lim_{n \to \infty} f(z_n) = f(\zeta),$$

i.e., that $f(\zeta) \in \mathscr{E}_1$ is a limit point of the *sequence* $\{f(z_n)\}$ of points in $\mathscr{E}_2$ and hence of the *set* $\mathscr{E}_2$.[3] Thus, if $\mathscr{E}$ is disconnected, so is E, and the theorem is proved.

THEOREM 4.3. *Every continuous curve L is a continuum.*

Proof. By definition, L is the continuous image of some closed interval $[a, b]$, i.e., of some segment of the real line. Since $[a, b]$ is compact and connected (see Theorem 4.1, Corollary), so is L, according to Theorems 3.15 and 4.2.

THEOREM 4.4. *Every arcwise connected set E is connected.*[4]

Proof. Suppose E is disconnected. Then there exist two nonempty disjoint sets E_1 and E_2 such that

$$E_1 \cup E_2 = E, \qquad E_1^d \cap E_2 = E_1 \cap E_2^d = 0.$$

Let z, z' be a pair of points with $z \in E_1$, $z' \in E_2$, and join z and z' by a continuous curve $L \subset E$. Moreover, let

$$L_1 = L \cap E_1, \qquad L_2 = L \cap E_2.$$

Then L_1 and L_2 are nonempty ($z \in L_1$, $z' \in L_2$) and disjoint. Since L is connected (see Theorem 4.3), at least one of the sets L_1 and L_2, say L_1, must contain a limit point, say ζ, of the other. But then ζ belongs to E_1 and is a limit point of E_2, contrary to hypothesis. This contradiction proves the theorem.

Example. The plane is an arcwise connected set, since any two of its points can be joined by a line segment (for example). Therefore, the plane is a connected set.

---

[3] Since none of the points $f(z_n)$ can coincide with $f(\zeta)$, the sequence $\{f(z_n)\}$ must contain infinitely many distinct points.

[4] We note that the converse is not true (cf. Problem 4.10). See however, Theorem 4.5.

16. Domains. Interior, Exterior and Boundary Points

An open connected set is called a *domain.*[5] According to Theorem 4.4, an open set is connected and hence is a domain, if it is arcwise connected. The converse is also true:

THEOREM 4.5. *Every domain G is arcwise connected.*

Proof. Suppose G is not arcwise connected, i.e., suppose there exist two points z' and z'' in G which cannot be joined by a continuous curve contained in G. Let G_1 denote the set of points in G which can be joined to z' by a continuous curve contained in G, and let G_2 denote the set consisting of all the remaining points of G. Obviously we have

$$G_1 \cup G_2 = G, \qquad G_1 \cap G_2 = 0.$$

Since $z' \in G$ and G is open, G contains a neighborhood $\mathcal{N}(z')$ of z'. Moreover, $\mathcal{N}(z') \subset G_1$, since every point of $\mathcal{N}(z')$ can be joined to its center z' by a line segment contained in G. Thus G_1 and G_2 are nonempty ($z'' \in G_2$).

If $z_1 \in G_1$, there exists a continuous curve $L \subset G$ joining z_1 to z'. Let $\mathcal{N}(z_1)$ be a neighborhood of z_1 such that $\mathcal{N}(z_1) \subset G$, and let $z \neq z_1$ be any point in $\mathcal{N}(z_1)$. Since z can be joined to z_1 by a line segment $\Delta \subset G$ (in fact, the one with end points z and z_1), z can be joined to z' by the continuous curve $L \cup \Delta$, and hence $z \in G_1$. Therefore $\mathcal{N}(z_1) \subset G_1$, which means that no point $z_1 \in G_1$ can be a limit point of G_2, i.e., $G_1 \cap G_2^d = 0$.

Finally, if $z_2 \in G_2$, let $\mathcal{N}(z_2)$ be a neighborhood of z_2 such that $\mathcal{N}(z_2) \subset G$, and let $z \neq z_2$ be any point in $\mathcal{N}(z_2)$. If z belonged to G_1, we could join z to z' by a continuous curve $L \subset G$. But then since z can be joined to z_2 by a line segment $\Delta \subset G$, we could also join z_2 to z' by the continuous curve $(L \cup \Delta) \subset G$, and hence z_2 would belong to G_1, contrary to hypothesis. Therefore $\mathcal{N}(z_2) \subset G_2$, which means that no point $z_2 \in G_2$ can be a limit point of G_1, i.e., $G_1^d \cap G_2 = 0$. Thus, we have decomposed G into two nonempty disjoint subsets G_1 and G_2, neither of which contains a limit point of the other. But this contradicts the assumption that G is connected, thereby completing the proof.

Remark. Because of Theorem 4.5, we can define a domain as an open, *arcwise* connected set.

Let O be an arbitrary open set, and let z_0 be any point of O. By the (*connected*) *component* of O containing the point z_0, we mean the set $\mathcal{C}(z_0)$ of all points in O which can be joined to z_0 by a continuous curve contained in O.

---

[5] A domain is often denoted by the letter G (from the German *Gebiet*). The concept of a domain as defined here should not be confused with the concept of the domain of definition of a function (see p. 24).

THEOREM 4.6. *Every component $\mathscr{C}(z_0)$ of an open set O is a domain.*

Proof. $\mathscr{C}(z_0)$ is nonempty, since it contains z_0 and some neighborhood of z_0. $\mathscr{C}(z_0)$ is open, since if z_1 can be joined to z_0 by a continuous curve $L \subset O$, then any point z in a neighborhood $\mathscr{N}(z_1)$ of z_1 such that $\mathscr{N}(z_1) \subset O$ can be joined to z_0 by a continuous curve contained in O, i.e., the union of L and the line segment with end points z and z_1. Finally, $\mathscr{C}(z_0)$ is arcwise connected, by definition.

Remark. $\mathscr{C}(z_0)$ can also be characterized as the largest domain contained in O and containing z_0.

It is obvious that two components of O with a point in common must coincide completely. If O is not a domain, i.e., if O is disconnected, then O has at least two different components. The set of components of O may be infinite, as in the case where O is the union of a family of open disks of radius $\leqslant \frac{1}{2}$ with centers at the points $0, \pm 1, \pm 2, \ldots$ (here the components of O are the separate disks). In any case, the set of components of O must be countable (cf. Problem 3.2). To see this, in each component of O we choose a point with rational coordinates x and y. This establishes a one-to-one correspondence between the set of all components of O and a subset of a countable set (the rational numbers), thereby proving our assertion. Clearly, O is the union of all its components. Thus we have proved

THEOREM 4.7. *Every open set is the union of countably many disjoint domains.*

Relative to any open set O (in particular, a domain) all the points of the plane fall into one of the following three categories:

1. Points which belong to O, called *interior points* of O;
2. Points which do not belong to O and are limit points of O, called *boundary points* of O;
3. Points which do not belong to O and are not limit points of O, called *exterior points of* O.

The set of all interior points of O is called the *interior* of O, the set of all boundary points of O is called the *boundary* of O, and the set of all exterior points of O is called the *exterior* of O. The interior of O coincides with O, and hence is nonempty whenever O is nonempty. If O is the whole plane, the boundary of O is empty. If O is not the whole plane, its complement O^c must be nonempty. Clearly, O^c can consist only of boundary points and exterior points. If O^c contains no exterior points, then O^c consists entirely of boundary points, of which there must obviously be at least one. Moreover, if O^c contains any exterior points at all, there must be an infinite number of such points (since every exterior point has a neighborhood all of whose points are also exterior points).

THEOREM 4.8. *Every open set O with exterior points has an infinite number of boundary points.*

Proof. Let z be an interior point of O, let z' be an exterior point of O, and let L be the line segment joining z to z'. Let L_1 be the set of all interior points of O belonging to L, and let L_2 be the set of all exterior points of O belonging to L; obviously L_1 and L_2 are nonempty and disjoint. Moreover, both $L_1^d \cap L_2$ and $L_1 \cap L_2^d$ are empty, since any interior point of O has a neighborhood consisting entirely of interior points, and similarly for any exterior point of O. However, $L_1 \cup L_2$ cannot equal L, since this would contradict the fact that L is connected. It follows that L contains at least one boundary point of O. To show that O has infinitely many boundary points, we draw the line segments joining z to all the points in some neighborhood $\mathcal{N}(z')$ consisting entirely of exterior points of O, and use the fact that each such segment contains at least one boundary point of O.

Remark. The above considerations show that unless an open set O is the whole plane, the boundary of O must be nonempty. However, O still does not need to have any exterior points. In fact, every exterior point of O is characterized by the fact that it has a whole neighborhood of points which do not belong to O. Therefore, if the set O^c consisting of the points not in O (i.e., the complement of O) contains no neighborhoods, O^c cannot contain any exterior points of O, and hence O^c consists entirely of boundary points of O. For example, this is the case if O consists of all points which do not belong to a given straight line or circle. Then the line or circle is the boundary of O, and O has no exterior points at all.

17. Simply and Multiply Connected Domains

THEOREM 4.9. *The boundary Γ of any open set O is a closed set.*

Proof. If $z \notin \Gamma$, then z must either be an interior point or an exterior point of O. In either case, there is a neighborhood of z containing no points of Γ, and hence $z \notin \Gamma^d$. Thus, no point not in Γ can be a limit point of Γ, and the theorem is proved.

By the *closure* $\bar{O}$ of an open set O, we mean the union of O and its boundary. It is clear that $\bar{O}$ is a closed set, since by definition, any limit point of O must belong to the interior or boundary of O. If O is a domain, the set $\bar{O}$ is called a *closed domain*. For example, the *closed disk*[6] with equation $|z - z_0| \leqslant \rho$ is a closed domain.

[6] Note the distinction between a *circle* C with equation $|z - z_0| = \rho$, and an open (or closed) *disk* K with equation $|z - z_0| < \rho$ (or $|z - z_0| \leqslant \rho$).

Let G be a bounded domain, so that every point of G satisfies the inequality $|z| < \rho$ for some ρ. Then all the points outside the circle $|z| = \rho$ are exterior points of G. It follows that the boundary points of G must satisfy the inequality $|z| \leqslant \rho$, i.e., the boundary Γ of a bounded domain is bounded (and hence compact, since Γ is closed). A bounded domain G is said to be *simply connected* if its boundary Γ is connected, and *multiply connected* if Γ is disconnected.

Example. The interior of any circle is a simply connected domain, while the set of points lying between two concentric circles is a multiply connected domain.

The class of multiply connected domains can be subdivided further: A simply connected domain is characterized by the fact that its boundary is a continuum. If the boundary of a multiply connected domain G can be represented as the union of a finite number $n > 1$ of disjoint continua, G is said to be *finitely connected*, specifically *n-connected*; otherwise, G is said to be *infinitely connected*. In stating this definition, we rely upon the last assertion of Problem 4.12 and the following

THEOREM 4.10. *Given a set E, suppose that*

$$E = F_1 \cup \cdots \cup F_n, \tag{4.6}$$

where $F_1, \ldots, F_n$ are n disjoint continua. Then the decomposition (4.6) *is unique.*

Proof. Suppose that in addition to (4.6), E has another decomposition

$$E = \Phi_1 \cup \cdots \cup \Phi_m$$

in terms of m disjoint continua $\Phi_1, \ldots, \Phi_m$. Then there must be a pair of sets Φ_j and F_k, where $\Phi_j \neq F_k$, such that $\Phi_{jk} = \Phi_j \cap F_k \neq 0$ and $\Psi_{jk} = \Phi_j \cap F_k^c \neq 0$. It is clear that

$$\Phi_{jk} \cup \Psi_{jk} = \Phi_j, \qquad \Phi_{jk} \cap \Psi_{jk} = 0,$$

and

$$\Psi_{jk} \subset (F_1 \cup \cdots \cup F_{k-1} \cup F_{k+1} \cup \cdots \cup F_n).$$

If ζ is a limit point of Φ_{jk}, then ζ is a limit point of both Φ_j and F_k, and hence $\zeta \in \Phi_{jk}$, $\zeta \notin \Psi_{jk}$. On the other hand, if ζ is a limit point of Ψ_{jk}, then ζ is a limit point of some F_l ($l \neq k$), and hence $\zeta \in F_l$, $\zeta \notin F_k$, $\zeta \notin \Phi_{jk}$. Thus, we have found a decomposition of the connected set Φ_j into two nonempty disjoint subsets Φ_{jk} and Ψ_{jk}, such that neither subset contains a limit point of the other. This contradiction proves the theorem.

Example 1. Let $K_1, \ldots, K_{n-1}$ be $n - 1$ disjoint closed disks, all contained in a given open disk K. Then the set

$$K - (K_1 \cup \cdots \cup K_{n-1})$$

is an n-connected domain. The boundary of K consists of n circles, i.e., the boundary C of the disk K and the boundaries C_j of the disks K_j $(j = 1, \ldots, n-1)$.

Example 2. Let K be the open unit disk $|z| < 1$, and delete from K its center $z = 0$ and all the closed disks

$$\left| z - \frac{1}{n} \right| \leqslant \frac{1}{2n(n+1)} \qquad (n = 2, 3, \ldots).$$

In this way, we obtain an infinitely connected domain whose boundary consists of the point $z = 0$ and the circles

$$|z| = 1, \qquad \left| z - \frac{1}{n} \right| = \frac{1}{2n(n+1)} \qquad (n = 2, 3, \ldots).$$

THEOREM 4.11. *Given a domain G with boundary Γ and a point z_0 in G, let $\delta = \rho(z_0, \Gamma)$.*[7] *Then G contains the neighborhood $\mathcal{N}(z_0; \delta)$.*

Proof. The assertion is obvious if $\Gamma = 0$, since in this case both G and $\mathcal{N}(z_0; \delta)$ are the whole plane. Thus suppose that $\Gamma \neq 0$. Then the neighborhood $\mathcal{N}(z_0; \delta)$ cannot contain points of Γ, since they all lie outside or on the circle $|z - z_0| = \delta$ (at least one point of Γ must belong to this circle). Moreover, $\mathcal{N}(z_0; \delta)$ cannot contain an exterior point z_1 of G, since the line segment joining z_0 to z_1 must contain at least one point of Γ. Therefore $\mathcal{N}(z_0; \delta)$ consists entirely of interior points of G, as asserted.

THEOREM 4.12. *Any two points of a domain G can be joined by a Jordan curve contained in G, which can always be chosen to be a polygonal curve.*

Proof. Let z and z' be two points of G. Then, by Theorem 4.5, there exists a continuous curve $L \subset G$ joining z to z', with equation

$$z = f(t) \qquad (a \leqslant t \leqslant b).$$

Let $\delta = \rho(L, \Gamma)$, where Γ is the boundary of G ($\delta > 0$ by Theorem 3.19). Since $f(t)$ is continuous on $[a, b]$, we can find points

$$a = t_0 < t_1 < \cdots < t_n = b$$

which are so close together that

$$|f(t_{j-1}) - f(t_j)| < \delta \qquad (j = 1, 2, \ldots, n).$$

Every disk $|z - f(t_j)| < \delta$ is contained in G, since the distance between $f(t_j)$ and Γ cannot be less than δ, and hence the line segment Δ_j joining

[7] If Γ is nonempty, $\delta > 0$ by Theorems 4.9 and 3.18. If Γ is empty, i.e., if G is the whole plane, we write $\delta = \infty$.

$f(t_{j-1})$ to $f(t_j)$, $j = 1, 2, \ldots, n$, is also contained in G. Therefore $\Lambda \subset G$, where Λ is the polygonal curve with segments $\Delta_1, \Delta_2, \ldots, \Delta_n$ which joins z to z'. If Λ does not intersect itself, Λ is a Jordan curve and the theorem is proved.

If Λ intersects itself, there exist two segments Δ_j and Δ_k $(1 \leqslant j \leqslant k)$ which have a common point z_0, where, if the segments are consecutive $(k = j + 1)$, z_0 can be taken to be different from the final point of Δ_j (which coincides with the initial point of Δ_{j+1}). If $k > j + 1$, let Δ_j^* be the part of Δ_j from z_0 to its final point, and let Δ_k^* be the part of Δ_k from its initial point to z_0. Then, by removing from Λ the closed polygonal curve

$$\Lambda^* = \Delta_j^* \cup \Delta_{j+1} \cup \cdots \cup \Delta_{k-1} \cup \Delta_k^*,$$

we obtain a new polygonal curve Λ_0, which still joins z to z', but has at least one segment less than Λ. If $k = j + 1$, one of the segments Δ_j and Δ_{j+1} which share a point z_0 (distinct from the common final point of Δ_j and initial point of Δ_{j+1}) must be part of the other segment. Suppose, for example, that $\Delta_j \subset \Delta_{j+1}$. Then, by removing from Λ the closed polygonal curve consisting of Δ_j and the part of the segment Δ_{j+1} from its initial point to the initial point of Δ_j, we obtain a new polygonal curve Λ_1, which still joins z to z', but has one or two segments less than Λ.

Thus, in every case where the original polygonal curve Λ intersects itself, we can replace Λ by another polygonal curve Λ_0 or Λ_1, which is still contained in G and joins z to z', but has at least one segment less than Λ. Repeating this operation of removing closed polygonal curves a finite number of times, we finally obtain a polygonal curve $\lambda \subset G$ (possibly consisting of only one segment) which does not intersect itself and joins z to z'. This completes the proof of the theorem.

18. The Jordan Curve Theorem

Given a closed set F, let F^c be its complement. Noting that F^c is open (see Problem 3.19), we consider the connected components of F^c (see p. 64). It is not hard to see that there may be infinitely many such components. For example, if F consists of the points belonging to the two infinite families of lines $x = 0, \pm 1, \pm 2, \ldots$ and $y = 0, \pm 1, \pm 2, \ldots$ parallel to the coordinate axes, F^c has infinitely many components, i.e., all open squares of the form

$$m < x < m + 1, \qquad l < y < l + 1 \qquad (m, n = 0, \pm 1, \pm 2, \ldots).$$

However, if F is a closed Jordan curve, F^c can only have two components, as shown by the following basic result:

THEOREM 4.13 (*Jordan curve theorem*).[8] *The complement* γ^c *of any closed Jordan curve* γ *has exactly two components, with* γ *as their common boundary. One of these components* $I(\gamma)$, *called the interior of* γ, *is bounded, and the other component* $E(\gamma)$, *called the exterior of* γ, *is unbounded.*

Example. The interior of the circle $|z - z_0| = \delta$ is the set of points satisfying $|z - z_0| < \delta$, and the exterior is the set of points satisfying $|z - z_0| > \delta$.

THEOREM 4.14. *A bounded domain* G *is simply connected if and only if whenever* G *contains a closed Jordan curve* γ, G *also contains* $I(\gamma)$.

Proof. First we prove that the condition is necessary. Let G be simply connected, and let γ be a closed Jordan curve contained in G. If $z \in \gamma$, there is a neighborhood $\mathcal{N}(z) \subset G$, since G is open. Since γ is the boundary of both $I(\gamma)$ and $E(\gamma)$, $\mathcal{N}(z)$ contains infinitely many points of both $I(\gamma)$ and $E(\gamma)$, and hence $G \cap I(\gamma) \neq 0$, $G \cap E(\gamma) \neq 0$. We cannot have $G \supset E(\gamma)$, since G is bounded (by the present definition of a simply connected domain), whereas $E(\gamma)$ is not (by Theorem 4.13). Therefore $E(\gamma)$ contains points of both G and its complement G^c. Moreover, $E(\gamma)$ contains points of Γ, where Γ is the boundary of G, since otherwise $E(\gamma)$ could be decomposed into two nonempty disjoint sets, i.e., the points of $E(\gamma)$ which are exterior points of G and the points of G which are interior points of G, neither of which contains a limit point of the other, contrary to the fact that $E(\gamma)$ is connected (being a domain). Now suppose G does not contain $I(\gamma)$. Then $I(\gamma)$ contains points of both G and G^c, and hence points of Γ, by the same argument as just given for $E(\gamma)$. It follows that Γ can be divided into two nonempty disjoint subsets, i.e., $\Gamma \cap I(\gamma)$ and $\Gamma \cap E(\gamma)$, neither of which contains a limit point of the other, contrary to the hypothesis that G is a simply connected domain. This contradiction shows that G contains $I(\gamma)$.

Next we prove that the condition is sufficient. Suppose G is bounded but not simply connected. Then we have to find a closed Jordan curve $\gamma \subset G$ such that G does not contain $I(\gamma)$. Since Γ is disconnected, there exist two nonempty disjoint sets Γ_1 and Γ_2 such that

$$\Gamma_1 \cup \Gamma_2 = \Gamma, \qquad \Gamma_1^d \cap \Gamma_2 = \Gamma_1 \cap \Gamma_2^d = 0. \tag{4.7}$$

Moreover, since Γ is closed (see Theorem 4.9), $\Gamma_1^d \subset \Gamma$, $\Gamma_2^d \subset \Gamma$ and therefore by (4.7), $\Gamma_1^d \subset \Gamma_1$, $\Gamma_2^d \subset \Gamma_2$, i.e., Γ_1 and Γ_2 are closed and

[8] The proof of Theorem 4.13 is beyond the scope of this book, but can be found in many textbooks on topology. See e.g., P. S. Aleksandrov, *Combinatorial Topology, Vol. 1* (translated by H. Komm), Graylock Press, Rochester, N. Y. (1956), Chap. 2. The concepts of the interior and exterior of a *closed Jordan curve* should not be confused with the unrelated concepts of the *interior* and *exterior* of a set (see p. 65 and Problem 4.14) which are denoted by different symbols.

hence compact (being bounded, like Γ itself). It follows from Theorem 3.19 that $\rho(\Gamma_1, \Gamma_2) > 0$. Let $\rho(\Gamma_1, \Gamma_2) = 4\varepsilon$, and draw the two infinite families of lines $x = 0, \pm\varepsilon, \pm 2\varepsilon, \ldots$ and $y = 0, \pm\varepsilon, \pm 2\varepsilon, \ldots$ parallel to the coordinate axes. These lines partition the z-plane into a family of closed squares of the form

$$m\varepsilon \leqslant x \leqslant (m+1)\varepsilon, \qquad n\varepsilon \leqslant y \leqslant (n+1)\varepsilon \qquad (m, n = 0, \pm 1, \pm 2, \ldots). \tag{4.8}$$

Let U denote the union of all the squares (4.8) which intersect Γ_1; clearly, there are only a finite number N of such squares, which we denote by

$$Q_1, Q_2, \ldots, Q_N. \tag{4.9}$$

Given two such squares, say Q_r and Q_s, we say that Q_r *is chained to* Q_s (or *vice versa*) if there exists a finite sequence

$$Q_r = Q_{i_1}, \qquad Q_{i_2}, \ldots, Q_{i_\nu} = Q_s \tag{4.10}$$

of squares (4.9), with Q_r as the first term and Q_s as the last term, such that every two consecutive squares in the sequence (4.10) have a side in common. By definition, any square is regarded as chained to itself. Then the N squares (4.9) can be grouped into k distinct classes ($1 \leqslant k \leqslant N$) such that every square in a given class is chained to every other square in the same class and to no square in any other class. Let A_j ($1 \leqslant j \leqslant k$) denote the closed domain which is the union of the squares in the jth class. Then U has the representation

$$U = A_1 \cup \cdots \cup A_k,$$

which is unique except for the order of the sets $A_1, \ldots, A_k$. It should be noted that $\Gamma_1 \subset U$ and $B \cap \Gamma_1 = B \cap \Gamma_2 = 0$, where B is the boundary of U (see Problem 4.15).

Now if A is any of the sets $A_1, \ldots, A_k$, then A contains at least one point of Γ_1. Moreover, the boundary of A consists of one or more closed polygonal Jordan curves $\pi_0, \pi_1, \ldots, \pi_l$, where one of the curves, say π_0, is such that $\overline{I(\pi_0)} \supset A$, and the other curves $\pi_1, \ldots, \pi_l$ are such that

$$\overline{E(\pi_1)} \supset A, \ldots, \overline{E(\pi_l)} \supset A.$$

If $E(\pi_0) \cap \Gamma_2 \neq 0$, then π_0 is a closed Jordan curve such that

$$\pi_0 \cap \Gamma = 0, \quad I(\pi_0) \cap \Gamma_1 \neq 0, \quad E(\pi_0) \cap \Gamma_2 \neq 0.$$

Otherwise, at least one of the curves $\pi_1, \ldots, \pi_l$, say π_1, is such that $I(\pi_1) \cap \Gamma_2 \neq 0$, and in this case π_1 is a closed Jordan curve such that

$$\pi_1 \cap \Gamma = 0, \quad I(\pi_1) \cap \Gamma_2 \neq 0, \quad E(\pi_1) \cap \Gamma_1 \neq 0.$$

Thus, in any event, we can find a closed Jordan curve γ (actually, a polygonal curve) such that

$$\gamma \cap \Gamma = 0, \quad I(\gamma) \cap \Gamma \neq 0, \quad E(\gamma) \cap \Gamma \neq 0.$$

To complete the proof, we show that G contains γ but not $I(\gamma)$. Since both $I(\gamma)$ and $E(\gamma)$ contain points of Γ, i.e., boundary points of G, both $I(\gamma)$ and $E(\gamma)$ contain points of G itself. Let $z \in I(\gamma)$ and $z' \in E(\gamma)$ be two points of G, and let $L \subset G$ be a continuous curve joining z to z' (such a curve exists, by Theorem 4.5). Then $L \cap \gamma \neq 0$ and hence γ contains at least one point of G. Since $\gamma \cap \Gamma = 0$, it follows that $\gamma \subset G$. But G cannot contain $I(\gamma)$, since, as already noted, $I(\gamma) \cap \Gamma \neq 0$ and hence $I(\gamma)$ contains at least one point not in G.

In the class of unbounded domains, to which we have not yet applied the concept of simple or multiple connectivity, there are domains which have the property of simply connected domains figuring in Theorem 4.14, i.e., domains which if they contain a closed Jordan curve γ also contain $I(\gamma)$. We shall call such domains simply connected, thereby extending the original concept of a simply connected domain to the case of unbounded domains.

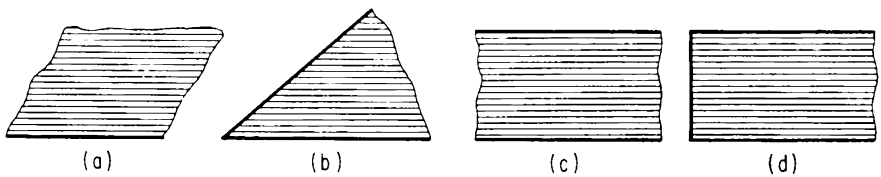

FIGURE 4.1

Example 1. The whole plane is simply connected, and so are the domains a, b, c and d shown in Figure 4.1, i.e., the half-plane, the wedge, the strip and the half-strip.

Example 2. The exterior of a circle is not simply connected, even with the extended definition. (The connectivity of domains of this kind will be discussed further on p. 85).

19. Some Further Results

THEOREM 4.15. *Let G be a domain, and let $F \subset G$ be a compact set. Then there exists another compact set $E \subset G$ and a number $\delta > 0$ such that E contains all the closed neighborhoods $\overline{\mathcal{N}(z; \delta)}$, $z \in F$, i.e., any compact set $F \subset G$ can be imbedded in another compact set $E \subset G$.*

Proof. First suppose that G is the whole plane. Then if every point $z \in F$ satisfies the inequality $|z| \leqslant \delta$, we can choose E to be the closed disk $|z| \leqslant 2\delta$, and all the requirements of the theorem are satisfied.

If G is not the whole plane, the boundary Γ of G is nonempty, and we can talk about the distance $d = \rho(F, \Gamma) > 0$ between F and Γ (see Theorem 3.19). Let E be the set of points z such that $\rho(z, F) \leqslant \frac{1}{2}d$. If $\delta = \frac{1}{2}d$, then E contains all the closed neighborhoods $\overline{\mathcal{N}(z; \delta)}$, $z \in F$,

and moreover $E \subset G$, according to Theorem 4.11. Since E is obviously bounded, it only remains to show that E is closed. Let ζ be a limit point of E, and let $\{z_n\}$ be a sequence of points in E converging to ζ. Since $\rho(z, F)$ is a continuous function of z (see the proof of Theorem 3.21), the sequence $\{\rho(z_n, F)\}$ converges to $\rho(\zeta, F)$. Moreover, by the construction of E, $\rho(z_n, F) \leqslant \frac{1}{2}d$, and hence $\rho(\zeta, F) \leqslant \frac{1}{2}d$. Therefore $\zeta \in E$, so that E is closed, as asserted. This proves the theorem, with $\delta = \frac{1}{2}d$.

THEOREM 4.16. *Let E be a nonempty open subset of a domain G, with the property that every limit point of E belonging to G also belongs to E. Then E coincides with the domain G.*

Proof. Suppose the theorem is false. Then there exist points of the domain G which do not belong to E. Let E_1 be the set consisting of these points. Since G is connected, at least one of the sets E and E_1 must contain a limit point of the other. However this set cannot be E, since E is open (and hence all points belonging to a sufficiently small neighborhood of a point in E also belong to E). Therefore E_1 must contain a limit point ζ of E, and by the hypothesis of the theorem ζ must belong to E. But this contradicts the fact that $\zeta \in E_1$ and thereby proves the theorem.

Remark. Let Γ be the boundary of G, let $\rho(z) = \rho(z, \Gamma)$, and suppose that $\mathcal{N}(z; \rho(z)) \subset E$ whenever $z \in E$. In this case, the conditions of Theorem 4.15 are satisfied. To see this, let $\zeta \in G$ be a limit point of E, and let $\rho(\zeta) = \rho(\zeta, \Gamma)$. Then for any point $z \in E$ in the neighborhood $\mathcal{N}(\zeta; \frac{1}{2}\rho(\zeta))$,[9] we have $\rho(z) \geqslant \frac{1}{2}\rho(\zeta)$, and hence $\mathcal{N}(z; \rho(z)) \subset E$ contains ζ. Thus every point $\zeta \in G$ which is a limit point of E belongs to E, as required. It follows that $E = G$.

The next theorem is of quite a different character:

THEOREM 4.17. *If F is the intersection of a sequence of continua $\{F_n\}$ such that $F_{n+1} \subset F_n$ $(n = 1, 2, \ldots)$, then F is also a continuum.*

Proof. Let $\{z_n\}$ be a sequence of points such that $z_n \in F_n$. Since $F_{n+1} \subset F_n$, the points $z_n, z_{n+1}, \ldots$ belong to F_n. It follows that all the limit points of $\{z_n\}$ belong to F_n $(n = 1, 2, \ldots)$ and hence to F. But the set of limit points of $\{z_n\}$ is nonempty, since $\{z_n\}$ is a bounded sequence. Therefore the set F is nonempty. If F has no limit points, it is a closed set (trivially). Suppose F has limit points, and let ζ be one of them. Since $F \subset F_n$, ζ is a limit point of F_n and hence belongs to F_n $(n = 1, 2, \ldots)$, i.e., $\zeta \in F$. Thus the set F is nonempty, closed and (obviously) bounded.

We still must show that F is connected. Suppose F is disconnected, so that F can be decomposed into two nonempty disjoint (compact) subsets D and E. Choose $z \in D$ and $z' \in E$. Since $z \in F_n$, $z' \in F_n$ for all n,

[9] At least one such point exists, since ζ is a limit point of E.

and since each F_n is a connected set, it follows from Theorem 4.1 that given any F_n, there exist points $z = z_0, z_1, \ldots, z_\nu = z'$ in F_n (which of course depend on the particular F_n) such that

$$|z_{k-1} - z_k| < \tfrac{1}{3}\rho(D, E) = \varepsilon \qquad (k = 1, 2, \ldots, \nu). \tag{4.11}$$

Among these points, there must be at least one point, say z_k, such that

$$\rho(z_k, D) > \varepsilon, \qquad \rho(z_k, E) > \varepsilon. \tag{4.12}$$

If this is not the case, then for each $k = 1, 2, \ldots, \nu$, either $\rho(z_k, D) \leqslant \varepsilon$ or $\rho(z_k, E) \leqslant \varepsilon$. Let D^* be the set of points z_k such that $\rho(z_k, D) \leqslant \varepsilon$, and let E^* be the set of points z_k such that $\rho(z_k, E) \leqslant \varepsilon$. Obviously, every point z_k $(k = 1, 2, \ldots, \nu)$ belongs to either D^* or E^*, and $\rho(D^*, E^*) \geqslant \varepsilon$, since $\varepsilon = \frac{1}{3}\rho(D, E)$. Let $\mu < \nu$ be the smallest subscript of the points in E^*. Then $z_{\mu-1} \in D^*$, and hence, according to (4.11),

$$|z_{\mu-1} - z_\mu| < \varepsilon,$$

which is impossible, since $\rho(D^*, E^*) \geqslant \varepsilon$. This contradiction proves the existence of the required point z_k satisfying (4.12). We henceforth denote z_k by ζ_n to emphasize that it belongs to F_n.

Thus $\zeta_n \in F_n$ and $\rho(\zeta_n, F) > \varepsilon$ for $n = 1, 2, \ldots$. Therefore, any limit point ζ of the (bounded) sequence $\{\zeta_n\}$ must satisfy the inequality $\rho(\zeta, F) \geqslant \varepsilon > 0$. But this is impossible, since all the limit points of $\{\zeta_n\}$ belong to F. It follows that F is a connected set, and the proof is complete.

PROBLEMS

4.1. If E is any set, prove that

a) E^d is closed;

b) $E^{dd} \subset E^d$.

Give an example of a set E such that $E^d \neq E^{dd} \neq E^{ddd} = 0$.

Hint. Consider the set of numbers of the form

$$\frac{1}{m} + \frac{1}{n} \qquad (m, n = 1, 2, \ldots).$$

4.2. Prove that if A and B are two disjoint closed subsets of a set E, and if a set $C \subset A \cup B$ is connected, then either $C \subset A$ or $C \subset B$.

4.3. Prove that if every two points of a set E belong to a connected subset of E, then E is connected.

4.4. A set E is said to be *convex* if whenever E contains two points z_1 and z_2, E also contains the line segment joining z_1 and z_2. Prove that every convex set is connected.

Hint. Use Problem 4.2 and Theorem 4.1, Corollary.

Comment. For further results on convex sets, see Problems 13.2–13.5.

4.5. Prove that if A and B are connected sets such that $A \cap B \neq 0$, then $A \cup B$ is connected.

4.6. Prove that the union of two disjoint open sets is disconnected.

4.7. A real function $\lambda(t)$ defined on a subset E of the real line is said to be *increasing* (or *nondecreasing*) on E if for every pair of points t' and t'' in E, $t' < t''$ implies $\lambda(t') \leqslant \lambda(t'')$. If $t' < t''$ implies $\lambda(t') < \lambda(t'')$, then $\lambda(t)$ is said to be *strictly increasing* on E. *Decreasing* (or *nonincreasing*) functions and *strictly decreasing* functions are defined similarly by writing $\lambda(t') \geqslant \lambda(t'')$ instead of $\lambda(t') \leqslant \lambda(t'')$ and $\lambda(t') > \lambda(t'')$ instead of $\lambda(t') < \lambda(t'')$. A function is said to be (*strictly*) *monotonic* on E if it is (strictly) increasing or (strictly) decreasing on E.

Prove that if $\lambda(t)$ is monotonic on an interval $[a, b]$, then $\lambda(t)$ has a left-hand limit $\lambda(t_0-)$ and a right-hand limit $\lambda(t_0+)$ at every point $t_0 \in (a, b)$, and

$$\lambda(t_0-) \leqslant \lambda(t_0) \leqslant \lambda(t_0+).$$

Morcover, prove that $\lambda(a+)$ and $\lambda(b-)$ exist, and

$$\lambda(a) \leqslant \lambda(a+), \qquad \lambda(b-) \leqslant \lambda(b).$$

4.8. Prove that if $\tau = \lambda(t)$ is strictly monotonic on an interval $[a, b]$, then the inverse function $t = \lambda^{-1}(\tau)$ exists and is strictly monotonic on its domain of definition.

4.9. Prove that the only connected subsets of the real line which contain more than one point are the intervals $a \leqslant x \leqslant b$, $a < x < b$, $a \leqslant x < b$ and $a < x \leqslant b$ (the values $a = -\infty$ and $b = +\infty$ are permitted). Prove that the only continua on the real line which contain more than one point are the closed intervals.

4.10. Prove that the set E of points (x, y) such that

$$y = \begin{cases} \sin \dfrac{1}{x} & \text{for} \quad 0 < x \leqslant 1, \\ 0 & \text{for} \quad -1 \leqslant x \leqslant 0 \end{cases}$$

is a continuum. Prove that E is not arcwise connected.

4.11. Prove the following result, often called the *intermediate value theorem*: Let $f(z)$ be a continuous real-valued function defined on a connected set E, and suppose that $f(z_1) = a$, $f(z_2) = b$, where z_1, z_2 are points of E and $a < b$. Then $f(z)$ takes any value between a and b at some point of E.

4.12. Let E be any set. By the (*connected*) *component* of E containing the point $z_0 \in E$, we mean the union of all connected subsets of E containing z_0. Prove that every point of E belongs to a uniquely determined component of E. Prove that this definition reduces to the one on p. 64 when E is open. Finally, prove that every component of a closed set is closed, and hence a continuum (if bounded).

4.13. By the *closure* of a set E, we mean the set $\bar{E} = E \cup E^d$. Prove that $\bar{E}$ is closed and that E is closed if and only if $E = \bar{E}$. Verify the relations

$$\overline{E \cup F} = \bar{E} \cup \bar{F}, \quad (\bar{E})^d = E^d, \quad \overline{(\bar{E})} = \bar{E}.$$

Prove that this definition reduces to the one on p. 66 when E is open.

4.14. By an *interior point* of a set E, we mean a point in E which has a neighborhood contained in E, and the set of all interior points of E is called the *interior* of E, denoted by int (E). Similarly, by an *exterior point* of a set E, we mean a point in E^c which has a neighborhood contained in E^c, and the set of all exterior points of E is called the *exterior* of E, denoted by ext (E). Prove that

a) int (E) and ext (E) are open;
b) int (E) = ext (E^c), ext (E) = int (E^c);
c) int $(E) = (E^c)^c$, ext $(E) = (\bar{E})^c$;
d) E is open if and only if E = int (E), while E is closed if and only if E^c = ext (E).

Comment. Obviously, these definitions reduce to the corresponding ones on p. 65 when E is open.

4.15. By a *boundary point* of a set E, we mean a point z such that every neighborhood of z contains at least one point of E and at least one point of E^c (z itself may belong to either E or E^c). The set of all boundary points of E is called the *boundary* of E, denoted by E^b. Prove that

a) E^b is closed (cf. Theorem 4.9);
b) $E^b = \bar{E} \cap \overline{E^c}$;
c) int (E), ext (E) and E^b are disjoint sets whose union is the whole complex plane.

Comment. Again, these definitions reduce to the corresponding ones on p. 65 when E is open.

4.16. Give an example of a domain D such that D and $\bar{D}$ have different boundaries.

4.17. Let E be the set consisting of all points of the upper half-plane Im $z > 0$ which do not lie on the family of line segments

$$z = \frac{1}{n} + it \qquad (0 \leqslant t \leqslant 1; n = \pm 1, \pm 2, \ldots).$$

What is the boundary of E? Is E a domain, and if so, is E simply connected?

5

INFINITY AND STEREOGRAPHIC PROJECTION

20. Proper and Improper Complex Numbers

In the theory of functions of a complex variable, an important role is played by an "improper" complex number called *infinity* and denoted by the symbol ∞. In order to define this concept suitably, we first introduce a new interpretation of complex numbers, which comes up quite naturally in problems involving limits.

Two convergent sequences of complex numbers $\{u_n\}$ and $\{v_n\}$ are said to be *coterminal* if they have the same limit.[1] Let E be the set of all convergent sequences of complex numbers, and divide E into all classes (subsets) of coterminal sequences. In other words, $\{u_n\}$ and $\{v_n\}$ belong to the same class if and only if $\{u_n\}$ and $\{v_n\}$ converge to the same limit. Obviously, these classes are disjoint and their union is E. If the common limit of all the sequences in a given class is denoted by a small letter, say a, then we denote the class itself by the corresponding capital letter, say A, i.e., $u_n \to a$ as $n \to \infty$ if and only if $\{u_n\} \in A$. In this way, we establish a one-to-one correspondence

$$a \leftrightarrow A, \qquad b \leftrightarrow B, \ldots \tag{5.1}$$

between the complex numbers and classes of coterminal sequences.

[1] For a somewhat different definition of coterminal sequences, which reduces to the present definition in the case where the sequences are convergent, see G. E. Shilov, *An Introduction to the Theory of Linear Spaces* (translated by R. A. Silverman), Prentice-Hall, Inc., Englewood Cliffs, N. J. (1961), p. 253.

If A and B are two classes of coterminal sequences, we define $A \pm B$ as the class of all sequences of the form $\{u_n \pm v_n\}$, AB as the class of all sequences of the form $\{u_n v_n\}$, and A/B as the class of all sequences of the form $\{u_n/v_n\}$, where $\{u_n\} \in A$, $\{v_n\} \in B$ but are otherwise arbitrary.[2] Since $\{u_n\} \in A$, $\{v_n\} \in B$, we have

$$\lim_{n \to \infty} u_n = a, \quad \lim_{n \to \infty} v_n = b$$

and hence

$$\lim_{n \to \infty} (u_n \pm v_n) = a \pm b, \quad \lim_{n \to \infty} u_n v_n = ab, \quad \lim_{n \to \infty} \frac{u_n}{v_n} = \frac{a}{b}. \tag{5.2}$$

It follows that $A \pm B$, AB and A/B are again classes of coterminal sequences. Moreover (5.2) implies that

$$a \pm b \leftrightarrow A \pm B, \quad ab \leftrightarrow AB, \quad \frac{a}{b} \leftrightarrow \frac{A}{B},$$

i.e., the one-to-one correspondence (5.1) is an *isomorphism*[3] between the field of complex numbers and the field of coterminal sequences. In this sense, we are justified in identifying a and A, b and B, etc., and then every class of coterminal sequences is called a *proper complex number*. In particular, the class A can be represented geometrically by the same point in the complex plane as that representing the complex number a. With this interpretation, a convergent sequence $\{u_n\}$ belongs to the class $A = a$ if and only if the unique limit point of the set of points representing the complex numbers $u_1, u_2, \ldots, u_n, \ldots$ coincides with a.

We now adjoin a single *improper complex number* to the set of all proper complex numbers (i.e., the set of all classes of coterminal sequences). This improper complex number, which we denote by ∞, is the class of all sequences $\{u_n\}$ with the property that given any $\rho > 0$, there exists an integer $n_0 > 0$ (depending on ρ and $\{u_n\}$) such that $|u_n| > \rho$ whenever $n > n_0$. If a sequence $\{u_n\}$ belongs to the class ∞, we say that $\{u_n\}$ *converges to infinity*, and we write $u_n \to \infty$ as $n \to \infty$ or

$$\lim_{n \to \infty} u_n = \infty.$$

The union of the set of all proper complex numbers and the improper complex number ∞ is called the extended complex number system. Algebraic operations are defined for the extended complex number system in exactly the same way as for the set of proper complex numbers. However, as the following examples show, in some cases an algebraic expression involving the class ∞ does not lead to a class of coterminal sequences, and hence is meaningless:

---

[2] Whenever we write u_n/v_n or a/b, it is assumed that $v_n \neq 0$ or $b \neq 0$.

[3] See e.g., G. Birkhoff and S. MacLane, *op. cit.*, p. 42.

1. If a is a proper complex number, then $a + \infty = \infty + a = \infty$, but $\infty + \infty$ is meaningless, as shown by the two sequences

$$1, 3, 3, 5, 5, 7, 7, 9, \ldots \tag{5.3}$$

and

$$-1, -2, -3, -4, -5, -6, -7, -8, \ldots \tag{5.4}$$

belonging to the class ∞. In fact, adding (5.3) and (5.4) term by term, we obtain the sequence

$$0, 1, 0, 1, 0, 1, 0, 1, \ldots,$$

which obviously does not belong to any class of coterminal sequences (since it has no limit).

2. Similarly, $a - \infty = \infty - a = \infty$ if a is a complex number, but $\infty - \infty$ is meaningless.
3. If $a \neq 0$ is a proper complex number, then $a \cdot \infty = \infty \cdot a = \infty$ and moreover $\infty \cdot \infty = \infty$, but $0 \cdot \infty$ is meaningless.
4. If a is any proper complex number, then $a/\infty = 0$ and $\infty/a = \infty$, but ∞/∞ is meaningless.

Remark. This approach also allows us to divide by zero. In fact, $a/0 = \infty$ if $a \neq 0$ is a proper complex number, but $0/0$ is meaningless.

21. Stereographic Projection. Sets of Points on the Riemann Sphere

In order to represent the extended complex number system geometrically, it is convenient to use the following construction, due to Riemann (1826–1866). Consider a sphere Σ of unit radius and center O, and let Π be a plane passing through O (see Figure 5.1). Introducing a rectangular coordinate system in the plane Π, with origin at O, we can represent any proper complex number $z = x + iy$ by a point (x, y) in the plane Π. To associate a point on the sphere Σ with a given point $P \in \Pi$, we first draw the diameter NS of the sphere which is perpendicular to Π and intersects Π at O. Then we draw the line segment joining one end of this diameter, say N, to the point P. The line segment NP (or its prolongation) intersects the sphere Σ in some point P^* different from N. It is clear that this construction establishes a one-to-one correspondence between the points of the sphere Σ (except for the point N itself) and the points of the plane Π. This mapping of the sphere into the plane (or of the plane into the sphere) is called *stereographic projection*, and the sphere Σ is called the *Riemann sphere*. If the $P \in \Pi$ represents the complex number z, we also regard the point $P^* \in \Sigma$ as representing z.

To be as descriptive as possible, we use geographic terminology. Thus, the circle in which the sphere Σ intersects the plane is called the *equator*, the points N and S are called the *north pole* and *south pole*, respectively, the great circles going through N and S are called *meridians*, and in particular, the meridian lying in the plane NOx is called the *prime* (or *initial*) *meridian*. Then it is easy to see that the points of the plane II lying inside the unit circle $|z| = 1$ (which coincides with the equator) are mapped into points of the *southern hemisphere* (containing S), while the points of II lying outside the unit circle are mapped into points of the *northern hemisphere* (containing N). Similarly, the upper half-plane $y > 0$ is mapped into the *eastern hemisphere* (which is intersected by the positive y-axis), while the lower half-plane $y < 0$ is mapped into the *western hemisphere*, and so on.

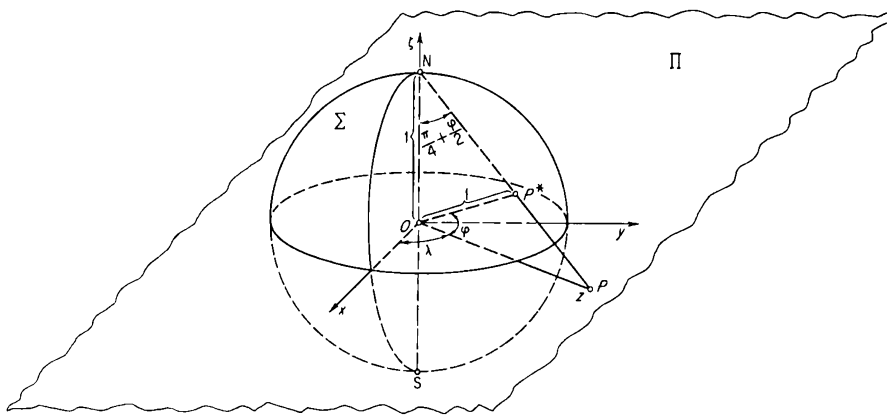

FIGURE 5.1

We now introduce spherical (or geographic) coordinates on Σ, i.e., the *latitude* φ, measured from the equator and ranging from 0 to $\pi/2$ in the northern hemisphere and from 0 to $-\pi/2$ in the southern hemisphere, and the *longitude* λ, measured from the prime meridian (more exactly, from the point of intersection of the prime meridian with the positive x-axis) and ranging from 0 to π (including π) in the eastern hemisphere and from 0 to $-\pi$ (excluding $-\pi$) in the western hemisphere. As shown by Figure 5.1, under stereographic projection we have

$$\arg z = \lambda, \qquad |z| = \tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right).$$

Therefore, the point of the sphere with coordinates λ and φ is the image of the complex number

$$z = \tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)(\cos\lambda + i\sin\lambda). \tag{5.5}$$

Conversely, the image on the Riemann sphere of the complex number $z \neq 0$ has spherical coordinates

$$\lambda = \arg z, \qquad \varphi = 2 \operatorname{arc\,tan} |z| - \frac{\pi}{2}.$$

Next we derive the formulas relating the rectangular coordinates (ξ, η, ζ) of the point P^* on the Riemann sphere and the rectangular coordinates $(x, y, 0)$ of the point $P \in \Pi$ which corresponds to P^* under stereographic projection. Since

$$\xi = \cos\varphi\cos\lambda, \quad \eta = \cos\varphi\sin\lambda, \quad \zeta = \sin\lambda,$$

and

$$\cos\varphi = \sin\left(\frac{\pi}{2} + \varphi\right) = \frac{2\tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)}{1 + \tan^2\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)} = \frac{2|z|}{1 + |z|^2},$$

$$\sin\varphi = -\cos\left(\frac{\pi}{2} + \varphi\right) = -\frac{1 - \tan^2\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)}{1 + \tan^2\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)} = \frac{|z|^2 - 1}{1 + |z|^2},$$

it follows that

$$\begin{aligned} \xi &= \cos\varphi\cos\lambda = \frac{|z|\cos\lambda}{1 + |z|^2} = \frac{2x}{x^2 + y^2 + 1}, \\ \eta &= \cos\varphi\sin\lambda = \frac{|z|\sin\lambda}{1 + |z|^2} = \frac{2y}{x^2 + y^2 + 1}, \\ \zeta &= \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1}. \end{aligned} \tag{5.6}$$

Moreover

$$z = \tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)(\cos\lambda + i\sin\lambda) = \frac{1 + \tan\frac{\varphi}{2}}{1 - \tan\frac{\varphi}{2}}(\cos\lambda + i\sin\lambda)$$

$$= \frac{\cos\frac{\varphi}{2} + \sin\frac{\varphi}{2}}{\cos\frac{\varphi}{2} - \sin\frac{\varphi}{2}}(\cos\lambda + i\sin\lambda) = \frac{\cos\varphi}{1 - \sin\varphi}(\cos\lambda + i\sin\lambda) = \frac{\xi + i\eta}{1 - \zeta},$$

and hence

$$x = \frac{\xi}{1 - \zeta}, \qquad y = \frac{\eta}{1 - \zeta}. \tag{5.7}$$

Using these formulas, we can prove the *circle-preserving property* of stereographic projection, i.e., we can show that stereographic projection maps circles on the Riemann sphere into circles or straight lines in the plane Π (the latter correspond to circles on Σ which go through the north pole N). To verify this, we first note that any circle on Σ is the intersection of Σ with some plane

$$A\xi + B\eta + C\zeta + D = 0. \tag{5.8}$$

It follows from (5.8) and (5.6) that the corresponding points of the plane satisfy the equation

$$Ax + By + (C + D)(x^2 + y^2) + (D - C) = 0,$$

which is the equation of a circle if $C + D \neq 0$, and the equation of a straight line if $C + D = 0$. But $C + D = 0$ is the result obtained when we substitute the coordinates (0, 0, 1) of the north pole into (5.8). Thus we obtain a circle or a straight line in the plane Π, depending on whether or not the circle on the sphere goes through the north pole N.

In particular, it should be noted that if the plane (5.8) is parallel to the equatorial plane $\zeta = 0$, i.e., if the circle on the sphere is a parallel of latitude, then $A = B = 0$, and in the plane Π we obtain a circle

$$(C + D)(x^2 + y^2) = C - D$$

with its center at the origin of coordinates. On the other hand, if the plane (5.8) goes through the axis NS, i.e., if the circle on the sphere is one of the meridians, then $C = D = 0$, and in the plane Π we obtain a straight line

$$Ax + By = 0$$

going through the origin of coordinates. Of course, all these results are apparent almost at once from the geometric meaning of stereographic projection.

Next we discuss sets of points on the Riemann sphere. Every circle γ on the sphere Σ which does not go through a given point $P^* \in \Sigma$ divides Σ into two parts, such that one part contains P^* and the other does not. The part of Σ containing P^* (and not including γ) will be called a *neighborhood* of P^* (cf. footnote 1, p. 27). Once the concept of the neighborhood of a point has been introduced, we can immediately extend the concepts of a limit point, a closed set, an open set, a continuous curve, a domain, etc., to sets of points on the surface Σ.

Example 1. A point $P^* \in \Sigma$ is a limit point of a given set $E \subset \Sigma$ if and only if any neighborhood of P^* contains infinitely many (distinct) points of E.

Example 2. Every triple of real functions

$$\xi = \varphi(t), \quad \eta = \psi(t), \quad \zeta = \chi(t)$$

of the real parameter t, defined and continuous on a closed interval $a \leqslant t \leqslant b$, and satisfying the condition

$$[\varphi(t)]^2 + [\psi(t)]^2 + [\chi(t)]^2 = 1,$$

defines a continuous curve on Σ.

Example 3. Let $\{a_n\}$ be a sequence of complex numbers converging to a (proper) complex number a. Let a_n $(n = 1, 2, \ldots)$ be represented by the point P_n in the plane Π and by the point P_n^* on the sphere Σ, and let a be represented by the point $P \in \Pi$ and by the point $P^* \in \Sigma$. Then P^* is the unique limit point of the sequence of points $\{P_n^*\}$, just as P is the unique limit point of the sequence of points $\{P_n\}$.

Now consider a sequence $\{a_n\}$ belonging to the improper class ∞ (see p. 78). Then, given any $\rho > 0$, there exists an integer $n > 0$ such that the points P_n representing the numbers a_n in the complex plane lie outside the circle $|z| = \rho$ if $n > n_0$. The corresponding points P_n^* of the Riemann sphere lie in the neighborhood of the north pole N bounded by the parallel of latitude whose projection onto the plane is the circle $|z| = \rho$. Therefore, the north pole is the unique limit point of every sequence belonging to the improper class. Conversely, every sequence of points on the sphere for which the north pole is the unique limit point is a sequence belonging to the improper class. This follows at once from the formula

$$|z| = \tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)$$

derived above, according to which the relations

$$\lim_{n \to \infty} |z_n| = +\infty$$

and

$$\lim_{n \to \infty} \varphi_n = \frac{\pi}{2}$$

are equivalent.[4] In keeping with this fact, *we shall henceforth regard the north pole N of the Riemann sphere as the geometric image of the improper complex number* ∞.

22. The Extended Complex Plane. The Point at Infinity

Let Σ denote the Riemann sphere, let $\Sigma - \{N\}$ denote Σ with the single point N (the north pole) deleted, and let Π be the ordinary or *finite* (complex) plane. Then, as we have just seen, stereographic projection is a one-to-one

[4] Note the distinction between the real number $+\infty$ and the complex number ∞. However, we usually write ∞ instead of $+\infty$ whenever $+\infty$ is a value of a real quantity which is inherently nonnegative, or when the context precludes any possibility of confusion.

mapping of $\Sigma - \{N\}$ onto Π which is continuous in both directions.[5] Therefore, $\Sigma - \{N\}$ is just as suitable as Π itself for representing the proper complex numbers. On the other hand, the presence on the sphere Σ of the extra point N, which it is natural to regard as the image of the complex number ∞, suggests that we think of Σ as the image of an *extended (complex) plane* $\Pi \cup \{\infty\}$, obtained by adjoining to the finite complex plane a single "improper" or "ideal" point, called the *point at infinity* and denoted by ∞, the same symbol as used to denote the corresponding improper complex number. In other words, by imagining that stereographic projection has been carried out, we can think of the extended plane as the union of the finite plane and a single extra point ∞, the point at infinity, which is the image under stereographic projection of the north pole of the Riemann sphere.

Remark. The extended plane should not be confused with the *projective plane*, which is obtained by adjoining to the finite plane an infinite set of improper points forming an improper or "ideal" line. To construct a model of the projective plane, we use a sphere Σ of unit radius which is *tangent* to a plane Π. Regarding the point of tangency as the *south pole* of Σ, we project the points of the southern hemisphere of Σ onto Π by drawing line segments from the *center* of Σ until they intersect both Σ and Π. This establishes a one-to-one correspondence between the points of the plane Π and the points of the southern hemisphere of Σ (except for the equator). The image of every line in Π is a great semicircle on Σ, and the image of a family of parallel lines in Π is a family of great semicircles on Σ, all of which pass through the same pair of diametrically opposite points of the equator (see Problem 5.8). This pair of points is regarded as the image of the *point at infinity* of the projective plane, common to all lines with the same direction. The set of all points at infinity of the projective plane corresponds to the set E of all pairs of diametrically opposite points of the equator. The set E is regarded as the image of a *line at infinity*. Thus the model of the projective plane is obtained from a hemisphere by identifying each pair of diametrically opposite points of the equator.

In the extended plane every sequence of points has at least one limit point. In fact, if the sequence is bounded, the existence of a (finite) limit point follows from Theorem 3.1 (the Bolzano-Weierstrass principle). If the sequence is unbounded, then the exterior of any circle, i.e., any neighborhood of the point at infinity, contains infinitely many terms of the sequence, and hence the point at infinity is the limit point of every unbounded sequence. Thus, although there are sequences in the ordinary or finite plane which

---

[5] A one-to-one mapping $P^* = f(P)$ of a set E onto a set $f(E)$ is said to be *continuous in both directions* if $f(P) \to f(P_0)$ as $P \to P_0$ and $f^{-1}(P^*) \to f^{-1}(P_0^*)$ as $P^* \to P_0^*$, for any $P_0 \in E$, $P_0^* = f(P_0) \in f(E)$.

have no limit points, *every sequence in the extended plane (or equivalently, on the Riemann sphere) has a limit point.* This fact is expressed by saying that the extended plane is *compact*, although the finite plane is not.[6]

As in the case of the finite plane, a sequence of points is said to be *convergent* if it has only one limit point, called the *limit* of the sequence, but now we allow for the possibility that the point at infinity is the limit of a sequence. Since the point at infinity is a limit point of any unbounded set E in the extended plane, E cannot be closed unless the point ∞ belongs to E. For example, the set of points $1, 2, \ldots, n, \ldots$, regarded as a set in the extended plane is not closed, (however, regarded as a set in the finite plane, it is closed).

If G is an unbounded domain, the point ∞ can either be a boundary point of G or an interior point of G. If ∞ is an interior point of G, there must exist a neighborhood of ∞ which belongs to G and hence contains no boundary points of G, i.e., in this case, the boundary of G must be a bounded set.

Example 1. The point at infinity is a boundary point of the half-plane, the wedge, the strip, the half-strip, and either of the two domains whose common boundary is a parabola or a branch of a hyperbola.

Example 2. The exterior G of any closed Jordan curve γ (e.g., a circle) has ∞ as an interior point, provided, of course, that G is regarded as containing ∞. Otherwise, we obtain another domain $G - \{\infty\}$, whose boundary consists of the curve γ and the isolated point at infinity.

In the extended plane, we can give a definition of the connectivity of a domain which does not depend on whether or not the domain is bounded, i.e., a domain is said to be *simply connected* if its boundary is a continuum,[7] and *multiply connected* otherwise. If the boundary of a multiply connected domain G can be represented as the union of a finite number of disjoint continua ($n > 1$), G is said to be *finitely connected*, specifically, *n-connected*; if not, G is said to be *infinitely connected.* These definitions agree with those previously given in Sec. 17 for the case of bounded domains. Moreover, the half-plane, the wedge, the strip, etc., are all simply connected domains, in keeping with the generalized definition given on p. 72, since in the extended plane the boundary of each of these unbounded domains is a continuum. However, in the case of a domain G which is the exterior of a closed Jordan curve γ, the old and new definitions do not agree, i.e., G is now a simply connected domain of the extended plane (if G is regarded as containing the point at infinity), even though G does not contain the interior of every closed Jordan curve γ whenever it contains γ itself. In fact, it can be shown

---

[6] Cf. Problem 3.23, which deals with the equivalence of various definitions of compactness.

[7] I.e., a closed connected set *on the Riemann sphere* (in the present context).

that with the present definition of a simply connected domain, the analogue of Theorem 4.14 is

THEOREM 5.1. *A domain G is simply connected if and only if whenever G contains a closed Jordan curve γ, G also contains either $I(\gamma)$ or $E(\gamma)$.*

Example. By deleting the point at infinity from the exterior of any closed Jordan curve, we obtain a doubly connected domain.

Having introduced the concept of the point at infinity, we now make a slight generalization of the concept of a function of a complex variable $f(z)$, by including ∞ in the set of possible values both of the independent variable z and of the function $f(z)$ itself. In other words, we allow $f(z)$ to take the value ∞ at a finite point z, and a finite or infinite value at the infinite point ∞. The appropriate way of doing this is to use the following continuity approach: If $f(z)$ is defined on a set E which has the point at infinity as a limit point, if $\{z_n\}$ is any sequence converging to ∞ (see p. 78), and if given any $\varepsilon > 0$, there exists an integer $n_0 > 0$ such that $|f(z_n) - c| < \varepsilon$ provided that $n > n_0$, where c is some finite complex number, then we write

$$\lim_{z \to \infty} f(z) = c,$$

and we set

$$f(\infty) = c$$

by definition. Similarly, if $f(z)$ is defined on a set E which has z_0 as a limit point (where z_0 may be finite or infinite), if $\{z_n\}$ is any sequence converging to z_0, and if given any $\rho > 0$, there exists an integer $n_0 > 0$ such that $|f(z_n)| > \rho$ provided that $n > n_0$, then we write

$$\lim_{z \to z_0} f(z) = \infty$$

and we set

$$f(z_0) = \infty$$

by definition. In each case, $f(z)$ is said to be (*wide-sense*) *continuous* at the point z_0 or ∞.

Example 1. If $f(z)$ is the rational function

$$f(z) = \frac{a_0 + a_1 z + \cdots + a_m z^m}{b_0 + b_1 z + \cdots + b_n z^n} \qquad (a_m \neq 0, \quad b_n \neq 0), \tag{5.9}$$

then

$$\lim_{z \to \infty} f(z) = \begin{cases} 0 & \text{for } m < n, \\ a_m/b_m & \text{for } m = n, \\ \infty & \text{for } m > n. \end{cases}$$

The function will be wide-sense continuous at ∞ if we set

$$f(\infty) = \lim_{z \to \infty} f(z)$$

in each case.

Example 2. Let z_0 be a *zero* of the polynomial $b_0 + b_1 z + \cdots + b_n z^n$ but not of the polynomial $a_0 + a_1 z + \cdots + a_m z^m$, i.e., suppose that

$$a_0 + a_1 z_0 + \cdots + a_m z_0^m \neq 0, \quad b_0 + b_1 z_0 + \cdots + b_n z_0^n = 0 \quad (a_m \neq 0, \quad b_n \neq 0).$$

Then the function $f(z)$ given by (5.9) approaches ∞ as $z \to z_0$, and $f(z)$ will be wide-sense continuous at z_0 if we set $f(z_0) = \infty$.

23. Conformality of Stereographic Projection. Continuous Curves in the Extended Plane

THEOREM 5.2. *Under stereographic projection, the angle between any two curves on the Riemann sphere* Σ *equals the angle between the images of the curves in the complex plane* Π, *and conversely.*

Proof. Let $P^* = (\xi, \eta, \zeta)$ be any point of the sphere other than the north pole, and let γ^* be a continuous curve passing through P^* which has a tangent at P^*. This means that among the various parametric representations of the curve γ^*, there is a representation

$$\xi = \varphi(t), \quad \eta = \psi(t), \quad \zeta = \chi(t) \qquad [\varphi(t)]^2 + [\psi(t)]^2 + [\chi(t)]^2 = 1,$$

such that at the parameter value corresponding to P^* the functions φ, χ, ψ are differentiable and[8]

$$[\varphi'(t)]^2 + [\psi'(t)]^2 + [\chi'(t)]^2 \neq 0.$$

Let the image of the point P^* under stereographic projection be the point $P = (x, y, 0)$, and let the image of the curve γ^* be the curve with parametric representation

$$x = \frac{\xi}{1 - \zeta} = \frac{\varphi(t)}{1 - \chi(t)}, \qquad y = \frac{\eta}{1 - \zeta} = \frac{\psi(t)}{1 - \chi(t)} \tag{5.10}$$

[sec (5.7)]. Obviously, γ passes through the point P. Differentiating (5.10), we find that the numbers $x'(t)$, $y'(t)$ characterizing the directions of the tangent at the point P are

$$x' = \frac{\xi'(1 - \zeta) + \zeta'\xi}{(1 - \zeta)^2}, \qquad y' = \frac{\eta'(1 - \zeta) + \zeta'\eta}{(1 - \zeta)^2},$$

so that

$$x'^2 + y'^2 = \frac{(\xi'^2 + \eta'^2)(1 - \zeta)^2 + 2(\xi\xi' + \eta\eta')\zeta'(1 - \zeta) + \zeta'^2(\xi^2 + \eta^2)}{(1 - \zeta)^4}.$$

[8] The prime denotes differentiation with respect to t.

However

$$\xi^2 + \eta^2 = 1 - \zeta^2, \qquad \xi\xi' + \eta\eta' = -\zeta\zeta', \tag{5.11}$$

and hence

$$x'^2 + y'^2 = \frac{(\xi'^2 + \eta'^2)(1-\zeta)^2 - \zeta'^2(2\zeta - 2\zeta^2) + \zeta'^2(1-\zeta^2)}{(1-\zeta)^4}$$
$$= \frac{(\xi'^2 + \eta'^2)(1-\zeta)^2 + \zeta'^2(1-\zeta)^2}{(1-\zeta)^4} = \frac{\xi'^2 + \eta'^2 + \zeta'^2}{(1-\zeta)^2}.$$

Now consider two curves γ_1^* and γ_2^* passing through the point P^* of the sphere, and let γ_1 and γ_2 be their images in the plane, passing through the point P. Then, if α is the angle between the curves γ_1 and γ_2 at the point P, we have

$$\cos\alpha = \frac{x_1'x_2' + y_1'y_2'}{\sqrt{x_1'^2 + y_1'^2}\,\sqrt{x_2'^2 + y_2'^2}}$$
$$= \frac{[\xi_1'(1-\zeta) + \zeta_1'\xi][\xi_2'(1-\zeta) + \zeta_2'\xi] + [\eta_1'(1-\zeta) + \zeta_1'\eta][\eta_2'(1-\zeta) + \zeta_2'\eta]}{(1-\zeta)^2\sqrt{\xi_1'^2 + \eta_1'^2 + \zeta_1'^2}\,\sqrt{\xi_2'^2 + \eta_2'^2 + \zeta_2'^2}}$$
$$= \frac{(\xi_1'\xi_2' + \eta_1'\eta_2')(1-\zeta)^2 + (\xi\xi_1' + \eta\eta_1')\zeta_2'(1-\zeta) + (\xi\xi_2' + \eta\eta_2')\zeta_1'(1-\zeta) + \zeta_1'\zeta_2'(\xi^2 + \eta^2)}{(1-\zeta)^2\sqrt{\xi_1'^2 + \eta_1'^2 + \zeta_1'^2}\,\sqrt{\xi_2'^2 + \eta_2'^2 + \zeta_2'^2}}.$$

But according to (5.11), generalized to the case of two curves γ_1^* and γ_2^*, we have

$$\xi^2 + \eta^2 = 1 - \zeta^2, \quad \xi\xi_1' + \eta\eta_1' = -\zeta\zeta_1', \quad \xi\xi_2' + \eta\eta_2' = -\zeta\zeta_2',$$

and hence

$$\cos\alpha = \frac{(\xi_1'\xi_2' + \eta_1'\eta_2')(1-\zeta)^2 - \zeta_1'\zeta_2'(\zeta - \zeta^2) - \zeta_1'\zeta_2'(\zeta - \zeta^2) + \zeta_1'\zeta_2'(1-\zeta^2)}{(1-\zeta)^2\sqrt{\xi_1'^2 + \eta_1'^2 + \zeta_1'^2}\,\sqrt{\xi_2'^2 + \eta_2'^2 + \zeta_2'^2}}$$
$$= \frac{\xi_1'\xi_2' + \eta_1'\eta_2' + \zeta_1'\zeta_2'}{\sqrt{\xi_1'^2 + \eta_1'^2 + \zeta_1'^2}\,\sqrt{\xi_2'^2 + \eta_2'^2 + \zeta_2'^2}} = \cos\alpha^*,$$

where α^* is obviously the angle between the curves γ_1^* and γ_2^* on the Riemann sphere at the point P^*. This completes the proof of the theorem.

Remark. A mapping of one surface (or domain on a surface) onto another is said to be *conformal* if it preserves angles between curves. Thus, according to Theorem 5.2, *stereographic projection of a sphere onto a plane is conformal.*

Applying Theorem 5.2, we can measure angles between curves on the Riemann sphere by measuring the angles between their images in the plane, and *vice versa.* We now use this approach to study the angles between

continuous curves passing through the point at infinity. First we introduce the concept of a continuous curve in the extended plane.

If the parametric equations of a continuous curve γ^* on the sphere are

$$\xi = \varphi(t), \quad \eta = \psi(t), \quad \zeta = \chi(t), \quad a \leqslant t \leqslant b,$$

then, according to (5.7), the equation of its image γ in the plane is

$$z = \frac{\xi + i\eta}{1 - \zeta} = \frac{\varphi(t) + i\psi(t)}{1 - \chi(t)} = f(t).$$

The function $f(t)$ is defined and continuous for all values of t in the interval $[a, b]$, with the exception of any value $t = t_0$ for which $\chi(t) = 1$. But when $\chi(t) = 1$, we have

$$[\varphi(t)]^2 + [\psi(t)]^2 = 1 - [\chi(t)]^2 = 0,$$

i.e., $\varphi(t) = \psi(t) = 0$. It follows that as $t \to t_0$, the curve γ^* approaches the north pole and hence $f(t) \to \infty$. Therefore, we can define $f(t)$ on the whole interval $[a, b]$ and preserve its continuity (in the wide sense) by setting $f(t_0) = \infty$. Guided by these considerations, we say that γ is a *continuous curve in the extended plane* if γ has an equation of the form $z = f(t)$, where $f(t)$ is wide-sense continuous on a closed interval $[a, b]$.[9] Obviously, if γ^* is the image of γ on the Riemann sphere Σ, then γ^* is a continuous curve on Σ, as previously defined (see p. 59), which may pass through the north pole.

The following are all examples of continuous curves in the extended plane:

1. The straight line

$$z = (at + b) + i(ct + d), \qquad -\infty < t < +\infty,$$

where $a^2 + c^2 \neq 0$ and $z = \infty$ for $t = \pm\infty$;

2. The parabola

$$z = (at^2 + bt + c) + i(dt + e), \qquad -\infty < t < +\infty,$$

where $a^2 + d^2 \neq 0$ and $z = \infty$ for $t = \pm\infty$;

3. The hyperbola

$$z = \frac{a(1 + t^2) + 2ibt}{1 - t^2}, \qquad -1 \leqslant t \leqslant 1,$$

where $z = \infty$ for $t = \pm 1$.

---

[9] We also allow the values $a = -\infty$, $b = +\infty$, provided that $f(t)$ approaches definite limits as $t \to a$, $t \to b$. However, by a continuous strictly monotonic change of parameter $\tau = \lambda(t)$ [see p. 60], we can transform $f(t)$, $a \leqslant t \leqslant b$, into another function $\varphi(\tau)$, $\alpha \leqslant \tau \leqslant \beta$, where $\varphi[\lambda(t)] \equiv f(t)$, and α, β are now finite. For example, under the parameter change $\tau = \lambda(t) = \arctan t$, $b = +\infty$ goes into $\beta = \pi/2$.

DEFINITION. *Two continuous curves γ_1 and γ_2 in the extended plane are said to form an angle of α radians with its vertex at infinity if their images γ_1^* and γ_2^* on the Riemann sphere under stereographic projection form an angle of α radians with its vertex at the north pole.*

Example 1. Consider two rays (half-lines) γ_1 and γ_2 which leave the point P at an angle α with each other. Their images γ_1^* and γ_2^* on the Riemann sphere are two circular arcs joining the point P^* (the image of P) with the north pole N. According to Theorem 5.2, the angle between γ_1^* and γ_2^* at P^* equals α. Moreover, by a familiar result of solid geometry (see Problem 5.11) the angle between γ_1^* and γ_2^* at N also equals α. It follows that the angle between γ_1 and γ_2 at ∞ equals α.

Example 2. Let γ_1 and γ_2 be two parallel lines. Their images γ_1^* and γ_2^* on the Riemann sphere are two circles which pass through N and have a common tangent at N. Therefore the angle between γ_1^* and γ_2^*, and hence between γ_1 and γ_2, equals 0.

Remark. The reader should become accustomed to thinking of the part of the extended plane between two half-lines as a figure with two equal angles and one vertex at ∞. Similarly, he should think of the part of the plane between two parallel lines (a strip) as a figure with two angles equal to zero and both vertices at ∞.

24. The Transformation $\zeta = 1/z$

Under stereographic projection, the simplest transformations of the Riemann sphere Σ into itself correspond to certain transformations of the extended plane into itself. For example, suppose we rotate Σ through the angle π about the diameter of Σ directed along the x-axis. Then the northern hemisphere goes into the southern hemisphere, and *vice versa*. In particular, the north and south poles (i.e., the points representing ∞ and 0) are interchanged. Moreover, the eastern and western hemispheres change places, while the equator and the prime meridian are transformed into themselves. It follows that the given transformation of the sphere into itself is equivalent to the transformation

$$\varphi \to -\varphi, \qquad \lambda \to -\lambda \tag{5.12}$$

of the spherical coordinates φ, λ of an arbitrary point of Σ.

To find the corresponding transformation of the extended plane into itself, we carry out the transformation (5.12) in formula (5.5). As a result, the point z goes into the new point

$$\zeta = \tan\left(\frac{\pi}{4} - \frac{\varphi}{2}\right)(\cos\lambda - i\sin\lambda) = \frac{1}{\tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)(\cos\lambda + i\sin\lambda)} = \frac{1}{z},$$

i.e., *rotation of the sphere through the angle π about the real axis corresponds to the transformation of the plane given by*

$$\zeta = \frac{1}{z}.$$

Under this transformation, the exterior of the unit circle (corresponding to the northern hemisphere) goes into the interior of the unit circle (corresponding to the southern hemisphere), and *vice versa*. Moreover, the upper and lower half-planes (corresponding to the eastern and western hemispheres) also change places, but, on the other hand, the unit circle (the image of the equator) and the real axis (the image of the prime meridian) transform into themselves.

It is particularly important to note that under the transformation $\zeta = 1/z$, the point at infinity and the origin of coordinates (the images of the north and south poles) change places, and furthermore, neighborhoods of these points change places. This is completely consistent with the relations

$$\infty = \frac{1}{0}, \qquad 0 = \frac{1}{\infty}$$

satisfied by the improper complex number ∞ (see Sec. 20). In fact, the transformation $\zeta = 1/z$ is habitually used in problems in which the point at infinity plays a role, since it reduces problems involving ∞ and its neighborhoods to problems involving 0 and its neighborhoods.

Remark 1. It follows from the fact that stereographic projection preserves circles (see Sec. 21) that the transformation $\zeta = 1/z$ also preserves circles. To see this, let γ be any circle or line in the z-plane, and let γ^* be the circle which is the image of γ on the sphere Σ under stereographic projection. Rotating Σ through π radians about the real axis carries γ^* into a new circle $\tilde{\gamma}^*$. But $\tilde{\gamma}^*$ is itself the image under stereographic projection of some circle or straight line $\tilde{\gamma}$ in the z-plane, and moreover $\tilde{\gamma}$ is just the image of γ under the transformation $\zeta = 1/z$. In other words, the transformation $\zeta = 1/z$, like stereographic projection itself, carries circles and straight lines into circles or straight lines.

Remark 2. In exactly the same way, we can prove that the transformation $\zeta = 1/z$ is conformal. In Secs. 33, 34 and 45, we shall prove without using stereographic projection that $\zeta = 1/z$ belongs to a large class of transformations which are conformal and preserve circles.

25. Another Definition of an Angle with its Vertex at Infinity

Instead of the definition of an angle with its vertex at infinity given in Sec. 23, we can give another, equivalent definition. First we recall that if

γ_1 and γ_2 are two continuous curves in the extended plane which form an angle of α radians with its vertex at infinity, then, according to the definition on p. 90, γ_1 and γ_2 are the images under stereographic projection of two continuous curves γ_1^* and γ_2^* on the Riemann sphere Σ, which form an angle of α radians with its vertex at the north pole. Rotating Σ through the angle π about the x-axis, we transform the curves γ_1^* and γ_2^* into two new curves $\tilde{\gamma}_1^*$ and $\tilde{\gamma}_2^*$, which form an angle of α radians with its vertex at the south pole. Let $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ be the images of $\tilde{\gamma}_1^*$ and $\tilde{\gamma}_2^*$ under stereographic projection. Then $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ form an angle of α radians with its vertex at the origin of coordinates (the image of the south pole). On the other hand, according to Sec. 24, the curves $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are the images of γ_1 and γ_2 under the transformation $\zeta = 1/z$. It follows that if two curves γ_1 and γ_2 of the extended plane form an angle of α radians with its vertex at infinity, then their images $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ under the transformation $\zeta = 1/z$ form an angle of α radians with its vertex at the origin.

Reversing this argument, we can easily convince ourselves that if $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ form an angle of α radians with its vertex at the origin, and if γ_1 and γ_2 are the images of $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ under the transformation $\zeta = 1/z$, then γ_1 and γ_2 form an angle of α radians with its vertex at infinity. Therefore, we can replace our original definition of an angle with its vertex at infinity by the following equivalent definition:

DEFINITION. *Two continuous curves γ_1 and γ_2 in the extended plane are said to form an angle of α radians with its vertex at infinity if their images $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ in the extended plane under the transformation $\zeta = 1/z$ form an angle of α radians with its vertex at the origin.*

PROBLEMS

5.1. Prove that if $\{u_n\}$ and $\{v_n\}$ are coterminal sequences, then

$$\lim_{n \to \infty} |u_n - v_n| = 0.$$

5.2. Give an example of an unbounded sequence which does not converge to infinity.

5.3. Suppose the sequence $\{z_n\}$ converges to infinity as $n \to \infty$. What does this imply about the behavior of Re z_n, Im z_n, $|z_n|$ and Arg z_n?

5.4. Give examples showing why $\infty - \infty$, $0 \cdot \infty$, ∞/∞ and $0/0$ are meaningless.

5.5. What is the relation satisfied by two points z_1 and z_2 which are the images under stereographic projection of a pair of diametrically opposite points of the Riemann sphere?

Ans. $z_1 \bar{z}_2 = -1$.

5.6. Derive a formula for $k(z_1, z_2)$, the length of the chord joining the images on the Riemann sphere of two points z_1 and z_2 in the complex plane. Examine the case where one of the points is the point at infinity.

Ans. $k(z_1, z_2) = \dfrac{2|z_1 - z_2|}{\sqrt{1 + |z_1|^2}\sqrt{1 + |z_2|^2}}$.

5.7. Define stereographic projection for the case where the Riemann sphere Σ is *tangent* to the complex plane Π, and find the relation between the coordinates of a point in Π and its image on Σ.

5.8. Let Σ and Π be the same as in the preceding problem. Find the relation between the coordinates of a point in Π and its image on Σ when the center of projection is the center of Σ instead of the north pole of Σ. Verify that the image of a family of parallel lines in Π all pass through the same pair of diametrically opposite points on the equator of Σ.

5.9. Prove Theorem 5.1 and its converse.

Hint. Cf. the proof of Theorem 4.14.

5.10. Give a purely geometric proof of Theorem 5.2.

5.11. Prove that the two angles of any figure formed by two circular arcs drawn on a sphere are equal.

5.12. Find the images of the following curves under the transformation $\zeta = 1/z$:

a) The family of circles $x^2 + y^2 = ax$;
b) The family of circles $x^2 + y^2 = by$;
c) The family of parallel lines $y = x + b$;
d) The family of lines $y = kx$ passing through the origin.

6

HOMEOMORPHISMS

26. The One-to-One Continuous Image of a Domain

Let $w = f(z)$ be a one-to-one continuous function mapping a set E onto a set $\mathscr{E}$. If the mapping is *continuous in both directions* (cf. footnote 5, p. 84), i.e., if $f(z)$ is continuous on E and its inverse function $f^{-1}(w)$ is continuous on $\mathscr{E}$, then the mapping is said to be a *homeomorphism* (synonymously, a *topological mapping* or a *topological transformation*). The set $\mathscr{E}$ is said to be the *homeomorphic image* of E, and E is said to be *homeomorphic* to $\mathscr{E}$ (and conversely).

THEOREM 6.1. *Let G be a domain, and let $f(z)$ be a one-to-one continuous function on G.*[1] *Then the set $f(G)$ is a domain, i.e., the one-to-one continuous image of a domain is a domain. Moreover, $f(z)$ is a homeomorphism.*

Proof.[2] Once we succeed in showing that the one-to-one continuous image of a domain is a domain, the proof that $f^{-1}(w)$ is continuous on $\mathscr{G} = f(G)$, and hence that $f(z)$ is a homeomorphism, is almost immediate. In fact, if w_0 is an arbitrary point of $\mathscr{G}$ and if z_0 is its image under $f^{-1}(w)$, we need only prove that given any neighborhood $\mathscr{N}(z_0)$, there is a neighborhood $\mathscr{N}(w_0)$ such that $f^{-1}[\mathscr{N}(w_0)] \subset \mathscr{N}(z_0)$. However, since $f[\mathscr{N}(z_0)]$ is a domain, it is sufficient to choose $\mathscr{N}(w_0)$ to be any neighborhood of w_0 such that $\mathscr{N}(w_0) \subset f[\mathscr{N}(z_0)]$.

[1] Here, and in the rest of this chapter, we have in mind domains in the extended plane and functions which are wide-sense continuous.

[2] The proof of Theorem 6.1 can be omitted on a first reading.

To show that $\mathcal{G}$ is a domain, we have to prove that $\mathcal{G}$ is both connected and open. The fact that $\mathcal{G}$ is connected is an immediate consequence of Theorem 4.2. We now prove that $\mathcal{G}$ is open. This part of the proof is rather complicated and makes repeated use of the Jordan curve theorem (Theorem 4.13).

Let w_0 be an arbitrary point of $\mathcal{G}$, with inverse image $z_0 = f^{-1}(w_0)$, and make the following choice of an open set $Q_0 \subset G$ containing z_0: If $z_0 \neq \infty$, let Q_0 be the interior of a square $d_0 \subset G$ with z_0 as its center, while if $z_0 = \infty$, let Q_0 be the exterior of a square $d_0 \subset G$ with the origin as its center (in either case, d_0 is the boundary of Q_0).[3] It can always be assumed that the sides of d_0 are parallel to the coordinate axes and do not pass through the point (if such exists) where $f(z)$ becomes infinite. Clearly, the image $\delta_0 = f(d_0)$ is a closed Jordan curve in the finite

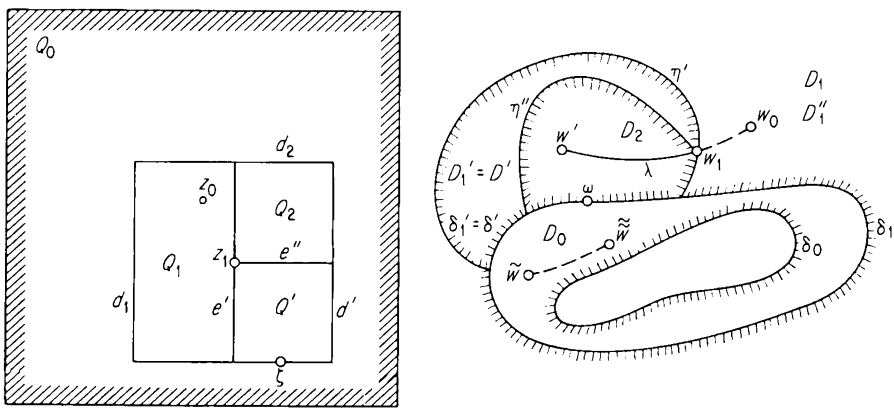

FIGURE 6.1

w-plane, with interior $I(\delta_0)$ and exterior $E(\delta_0)$. Moreover, $w_0 \notin \delta_0$, since $z_0 \notin d_0$, and hence w_0 belongs to either $I(\delta_0)$ or $E(\delta_0)$. Our plan is to show that $f(Q_0) = I(\delta_0)$ if $w_0 \in I(\delta_0)$, while $f(Q_0) = E(\delta_0)$ if $w_0 \in E(\delta_0)$. Then, since $f(Q_0) \subset \mathcal{G}$, it will follow in either case that w_0 belongs to an open set contained in $\mathcal{G}$, and hence that $\mathcal{G}$ is open, since w_0 is arbitrary.

Now, of the two domains $I(\delta_0)$ and $E(\delta_0)$ with boundary δ_0, let D_0 be the domain containing w_0 (see Figure 6.1, where the shading indicates *interiors* of closed Jordan curves). It is clear that $f(Q_0) \subset D_0$, since if $z \in Q_0$ and $l \subset Q_0$ is a Jordan curve joining z_0 to z (such a curve exists by Theorem 4.12), the set $f(l)$ is a Jordan curve joining $w_0 = f(z_0)$ to

[3] Here, by a *square* with *center* (a, b), we mean the closed Jordan curve consisting of the points (x, y) such that $x = a \pm c$, $b - c \leqslant y \leqslant b + c$ or $a - c \leqslant x \leqslant a + c$, $y = b \pm c$ $(c > 0)$.

$w = f(z)$. But $f(l) \cap \delta_0 = 0$ since $l \cap d_0 = 0$, and therefore, since the initial point w_0 of $f(l)$ belongs to D_0, so must its final point w. Thus, to prove that $f(Q_0) = D_0$, we must prove that $f(Q_0) \supset D_0$, i.e., that every point in D_0 is the image of some point in Q_0. We now show that the assumption that the opposite is true, i.e., that there is a point $w' \in D_0$ which does not belong to $f(Q_0)$, eventually leads to a contradiction, thereby completing the proof of the theorem.

Let $\lambda_0 \subset D_0$ be a Jordan curve joining w_0 to w'. Then the set

$$E = f(Q_0) \cap \lambda_0$$

is closed. In fact, let α be a limit point of E. Since λ_0 is closed (being a continuum), α must belong to λ_0. Let w_n be a sequence of points in E converging to α. The sequence of inverse images $\{z_n\}$ of these points belongs to Q_0, and hence $\{z_n\}$ contains a subsequence $\{z_{k_n}\}$ converging to some point a in $\bar{Q}_0$ (the closure of Q_0). Since $f(z)$ is continuous on $\bar{Q}_0$,

$$\lim_{n \to \infty} w_{k_n} = \lim_{n \to \infty} f(z_{k_n}) = f(a),$$

while on the other hand,

$$\lim_{n \to \infty} w_{k_n} = \alpha.$$

It follows that $\alpha = f(a)$. However, $a \notin d_0$, since $\alpha \notin \delta_0$. Therefore $a \in Q_0$, so that $\alpha = f(a) \in f(Q_0)$. Moreover, since $\alpha \in \lambda_0$, as just noted, $\alpha \in f(Q_0) \cap \lambda_0 = E$, i.e., E is closed, as asserted.

Since E is closed, moving along λ_0 from the point $w_0 \in E$ to the point $w' \notin E$, we must encounter for the last time a point belonging to E. In fact, let $w = \varphi(t)$ be the equation of λ_0, and let

$$w_0 = \varphi(t_0), \qquad w' = \varphi(t'),$$

where $t_0 < t'$. Then the values of the parameter t such that $\varphi(t) \in E$ have a least upper bound τ. Obviously $\tau < t'$, since $\varphi(t') \notin E$. Moreover $\varphi(\tau) \in E$, since the function $\varphi(t)$ is continuous and the set E is closed. Therefore, the last point of the set E encountered on the way from w_0 to w' is precisely $w_1 = \varphi(\tau)$. In this way, we arrive at a new Jordan curve $\lambda \subset \lambda_0$, i.e., the part of λ_0 going from w_1 to w', such that its initial point w_1 belongs to $f(Q_0)$ while all the rest of its points do not.

Let z_1 be the inverse image of w_1. We now construct a new square d_1, which bears the same relation to the point z_1 in the domain Q_0 as the square d_0 bears to the point z_0 in the domain G. Let Q_1 be the interior of d_1, and consider the image $\delta_1 = f(d_1)$ which is a closed Jordan curve in the finite w-plane. Of the two domains with boundary δ_1, let D_1 be the domain containing w_1 and the whole image $f(Q_1)$ of the domain Q_1 (see Figure 6.1). Since $w_1 = f(z_1)$ is the only point that λ shares with $f(Q_0)$, and since $w_1 \notin \delta_1$ ($z_1 \notin d_1$), λ shares no points with $\delta_1 \subset f(Q_0)$, and hence

λ must lie entirely in one of the two domains with boundary δ_1, i.e., in the domain D_1 containing the initial point w_1 of the curve λ.

Next we divide Q_1 into two congruent rectangular subdomains, by drawing the straight line segment e' joining the midpoints of the horizontal sides of d_1. The image $f(e')$ of the segment e' is a Jordan curve η' (which is not closed) with end points belonging to δ_1 and all other points belonging to D_1. The end points of η' divide δ_1 into two Jordan curves (whose union is δ_1), which together with η' form two closed Jordan curves δ_1' and δ_1'', each serving as the boundary of two domains. Of the two domains with boundary δ_1', let $\tilde{D}_1$ be the domain containing some exterior point $\tilde{w}$ of D_1. Then D_1', the other domain with boundary δ_1', is included in D_1. In fact, suppose the opposite is true, i.e., suppose D_1' contains some point $w_1' \notin D_1$. Then w_1' is either an exterior point or a boundary point of D_1. In the latter case, $w_1' \in \delta_1$ and hence w_1' is a boundary point of the exterior of D_1, i.e., any neighborhood $\mathcal{N}(w_1') \subset D_1'$ contains exterior points of D_1. Thus, in either case, D_1' contains exterior points of D_1. Let $\tilde{\tilde{w}}$ be such a point, and join $\tilde{\tilde{w}}$ to the point $\tilde{w}$ (which is also an exterior point of D_1) by a continuous curve contained in the exterior of D_1. Then this curve does not intersect either δ_1 or η', and hence does not intersect the boundary δ_1' of the domain D_1'. But this is impossible, since $\tilde{\tilde{w}} \in D_1'$ and $\tilde{w} \notin D_1'$. This contradiction proves that $D_1' \subset D_1$. In just the same way, of the two domains with boundary δ_1'', let $\tilde{D}_1''$ be the domain containing $\tilde{w}$. Then D_1'', the other domain with boundary δ_1'', is included in D_1.

It follows that every exterior point of D_1 belongs to either $\tilde{D}_1'$ or $\tilde{D}_1''$, and hence to the open set $O = \tilde{D}_1' \cap \tilde{D}_1''$. Moreover, no point of the curve δ_1 can belong to O, since any such point is a boundary point of either $\tilde{D}_1'$ or $\tilde{D}_1''$. Therefore, since O contains (all the) exterior points of D_1 and no boundary points of D_1, O can contain no interior points of D_1 (cf. the proof of Theorem 4.8), i.e., O coincides with the exterior of D_1. But this implies that any point $w \in D_1$, $w \notin \eta'$ must belong to either D_1' or D_1'', since otherwise $w \in \tilde{D}_1' \cap \tilde{D}_1'' = O$ $(w \notin \delta_1', \ w \notin \delta_1'')$, which is impossible $(D_1 \cap O = 0)$. Thus, one of the two domains D_1' and D_1'' contains the point w'. Let D' denote this domain, and let δ' be its boundary (δ' is one of the two curves δ_1', δ_1''). Then $\lambda - \{w_1\} \subset D'$, since $\lambda - \{w_1\} \subset D_1$, $\lambda \cap \eta' = \{w_1\}$. Let ω be a boundary point of the domain D', such that ω belongs to δ_1 and is different from the end points of η', and let ζ be the inverse image of ω. Then $\zeta \in d_1$ and ζ is different from the end points of e'. If $\{\zeta_n\}$ is a sequence of points in the domain Q_1 such that $\zeta_n \notin e'$ and $\zeta_n \to \zeta$ as $n \to \infty$, then the sequence $\{\omega_n\}$, where $\omega_n = f(\zeta_n)$, is contained in the domain D_1 and $\omega_n \to \omega$ as $n \to \infty$. It follows that the domain D' contains points of the sequence $\{\omega_n\}$, since otherwise, all the points ω_n would belong to the domain D_1' (or D_1'') which differs from D', and hence their limit point ω would belong to

the closure of this domain, which is impossible, because of our choice of ω. If $\omega_{n_0} = f(\zeta_{n_0})$ belongs to D' and if the point ζ_{n_0} belongs to the domain Q' (where Q' is one of the two domains into which the segment e' divides Q_1), then $f(Q') \subset D'$.

In this way, instead of the domain Q_1, whose image $f(Q_1)$ is included in the domain D_1, with boundary $\delta_1 = f(d_1)$, we obtain a new domain $Q' \subset Q_1$ whose boundary is a rectangle d' (see Figure 6.1), such that $f(Q')$ is included in the domain $D' \subset D_1$, with boundary $\delta' = f(d')$. It is essential to note that both the domain D_1 and the domain D' contain the fixed curve $\lambda - \{w_1\}$. If we divide the domain Q' into two congruent square subdomains by drawing the line segment e'' joining the midpoints of the vertical sides of d', then, by the same argument as before, we can select one of these domains, say Q_2, with boundary d_2, such that

$$f(Q_2) \subset D_2 \subset D_1, \quad w' \in D_2, \quad \lambda - \{w_1\} \subset D_2,$$

where D_2 is one of the two domains which have the closed Jordan curve $\delta_2 = f(d_2)$ as their common boundary. Here δ_2 is made up of parts of the three curves η', δ_1, $\eta'' = f(e'')$, and passes through the point w_1, since e' and e'' intersect at the point z_1 (which is the center of d_1 and the inverse image of w_1).

By repeated application of this argument, we arrive at a sequence

$$Q_0 \supset Q_1 \supset Q_2 \supset \cdots \supset Q_n \supset \cdots$$

of domains in the z-plane, each strictly including the next (cf. the proof of Theorem 3.1), whose boundaries are certain squares $d_0, d_1, d_2, \ldots, d_n, \ldots$ At the same time, we find a sequence of domains

$$D_0 \supset D_1 \supset D_2 \supset \cdots \supset D_n \supset \cdots$$

in the w-plane, each strictly including the next, such that

$$D_n^b = \delta_n = f(d_n), \quad f(Q_n) \subset D_n, \quad \lambda - \{w_1\} \subset D_n \quad (n = 0, 1, 2, \ldots),$$

where D_n^b denotes the boundary of D_n. Furthermore, there is a unique point z_1 in all the closed domains $\bar{Q}_n$, and hence all the domains $\bar{Q}_n$ must eventually (i.e., starting from some value of n) lie in an arbitrarily small neighborhood of z_1. Since $f(z)$ is continuous at z_1, it follows that all the closed Jordan curves δ_n must eventually lie in an arbitrarily small neighborhood of the point $w_1 = f(z_1)$. But this is prevented by the fact that all the domains D_n contain the fixed curve $l - \{w_1\}$, with end point $w' \neq w_1$.[4] This contradiction shows that there is no point $w' \in D_0$ which does not belong to $f(Q_0)$. Therefore $f(Q_0) \supset D_0$, and the proof of the theorem is complete!

---

[4] Here we have made essential use of the fact that each domain D_n contains the next. This can be seen most easily by using stereographic projection to visualize the whole construction as taking place on the Riemann sphere rather than in the extended plane.

Remark 1. Any homeomorphism is a one-to-one continuous function (in fact, continuous in both directions). Therefore, the homeomorphic image of a domain is a domain (see Problem 6.9).

Remark 2. We have not used the connectedness of G in proving that G is open. Therefore, the one-to-one continuous image of an open set is open.

27. Some Further Results

THEOREM 6.2. *Let G be a domain with boundary γ, let $f(z)$ be a one-to-one function on G, and let $f(z)$ be continuous on $\overline{G}$. Then $f(\gamma) \cap f(G) = 0$, i.e., none of the values of $f(z)$ on the boundary of G is attained at an interior point of G.*

Proof. Suppose the opposite is true, i.e., suppose

$$f(\zeta) = f(z_0) = w_0 \quad \text{where} \quad \zeta \in \gamma,\ z_0 \in G. \tag{6.1}$$

Consider a sequence $\{z_n\}$, $z_n \in G$ such that $z_n \to \zeta$ as $n \to \infty$. Then, since $f(z)$ is continuous on G, $w_n = f(z_n) \to f(\zeta) = w_0$. Let $\mathcal{N}(z_0) \subset G$ be a neighborhood of the point z_0 which does not contain any points of z_n, starting from some value of n. According to Theorem 6.1, $f[\mathcal{N}(z_0)]$ is a domain, and according to (6.1), $w_0 \in f[\mathcal{N}(z_0)]$. Therefore all the points of w_n, starting from some value of n, belong to $f[\mathcal{N}(z_0)]$, and therefore their inverse images z_n must belong to $\mathcal{N}(z_0)$. But this contradicts the way $\mathcal{N}(z_0)$ has been chosen, thereby proving the theorem.

THEOREM 6.3. *Let G be a domain with boundary γ, let $f(z)$ be a one-to-one function on G, and let $f(z)$ be continuous on $\overline{G}$. Then $f(\gamma) = \Gamma$, where Γ is the boundary of the domain $f(G)$, i.e., the image of the boundary of G coincides with the boundary of the image of G.*

Proof. Let $w \in f(\gamma)$, so that $w = f(z)$, where $z \in \gamma$. Consider a sequence of points $\{z_n\}$, $z_n \in G$, such that $z_n \to z$ as $n \to \infty$. Then $w_n = f(z_n) \to w$ as $n \to \infty$, where, of course, $w_n \in f(G)$. However, according to Theorem 6.2, $w = f(z) \notin f(G)$. Therefore w is a limit point of $f(G)$ which does not belong to $f(G)$, i.e., $w \in \Gamma$. It follows that $f(\gamma) \subset \Gamma$, since w is an arbitrary point of $f(\gamma)$.

Conversely, suppose $w \in \Gamma$. Then there exists a sequence $\{w_n\}$ such that $w_n \in f(G)$ and $w_n \to w$ as $n \to \infty$. If z_n is the inverse image of w_n, then $z_n \in G$, and from the sequence $\{z_n\}$ we can select a subsequence $\{z_{k_n}\}$ converging to a point $z \in \overline{G}$, where

$$f(z) = \lim_{n \to \infty} f(z_{k_n}) = \lim_{n \to \infty} w_{k_n} = w.$$

Thus w is the image of a point $z \in \overline{G}$. But $z \notin G$, since, according to Theorem 6.1, the image of an interior point of G is an interior point of

$f(G)$. Therefore z is a boundary point of G and $w = f(z)$ belongs to $f(\gamma)$. It follows that $f(\gamma) \supset \Gamma$, and the proof is complete.

Remark. Suppose that in the w-plane we know the image $\Gamma = f(\gamma)$ of the boundary γ of a domain G. Then to find $f(G)$, the image of G itself, we need only examine those components of Γ^c whose boundaries coincide with Γ. It may turn out that there is more than one such component. In this case, we can choose any point $z_0 \in G$ and determine which component of Γ^c contains $f(z_0)$. This component will then be the required image of G.

THEOREM 6.4. *The homeomorphic image of a simply connected domain G is a simply connected domain.*

Proof. Let C be an arbitrary closed Jordan curve contained in $f(G)$, and let $c = f^{-1}(C)$ be the inverse image of C. Obviously, c is a closed Jordan curve contained in G. Since G is simply connected, it follows from Theorem 5.1 that G also contains either $I(c)$ or $E(c)$, say $I(c)$. Then, according to Theorems 6.1 and 6.3, $f[I(c)]$ is a domain with boundary C. But $f(G)$ obviously contains $f[I(c)]$, since G contains $I(c)$. The same argument applies to $E(c)$ if $E(c) \subset G$. Therefore, using Theorem 5.1 again, we see that $f(G)$ is simply connected.

COROLLARY. *The homeomorphic image of a multiply connected domain G is a multiply connected domain.*

Proof. Note that $f^{-1}(w)$ is continuous on $f(G)$.

PROBLEMS

6.1. Find a homeomorphic mapping of the open interval $(-\pi/2, \pi/2)$ onto the whole real line.

6.2. Give an example of a one-to-one continuous mapping which is not a homeomorphism.

6.3. In the proof of Theorem 6.1, why is Q_1 always the interior of d_1?

6.4. In the proof of Theorem 6.1, show that no further special argument is needed to locate the domains $Q_3, Q_4, \ldots$

6.5. Prove that if F is a compact set and if $f(z)$ is a one-to-one continuous function on F, then $f(z)$ is a homeomorphism.

Hint. Use Theorem 3.15.

6.6. Prove that a Jordan curve is the homeomorphic image of a closed interval, and that a closed Jordan curve is the homeomorphic image of a circle.

6.7. Prove that a continuous real function $f(t)$ defined on a closed interval $a \leqslant t \leqslant b$ is a homeomorphism if and only if $f(t)$ is strictly monotonic (cf. Problem 4.7).

6.8. Without recourse to Theorem 6.1, give an elementary proof of the following fact: A one-to-one continuous mapping of the extended plane *onto* itself is a homeomorphism.

6.9. Without recourse to Theorem 6.1, prove that the homeomorphic image of a domain is a domain.

Hint. First show that if $f(z)$ is continuous and if E is an open set contained in the range of $f(z)$, then $f^{-1}(E)$ is open.

6.10. Prove the following analogue of Theorem 6.1: If E is an open set on the real line $-\infty < t < \infty$ and if $f(t)$ is a one-to-one continuous real function on E, then $f(E)$ is open.

Hint. Use the fact that every open set on the real line is the union of countably many open intervals.

6.11. Let γ be a (continuous) Jordan curve. Prove that the interior of γ is empty (cf. Problem 4.14).

Comment. The fact that γ is a Jordan curve is essential.

Part 2

DIFFERENTIATION ELEMENTARY FUNCTIONS

7

DIFFERENTIATION AND THE CAUCHY-RIEMANN EQUATIONS

28. Derivatives and Differentials

Let $f(z)$ be a function of a complex variable which is defined and single-valued on a set E, and let z_0 be any point of E which is a limit point of E. Then the *difference quotient*

$$\frac{f(z) - f(z_0)}{z - z_0} \tag{7.1}$$

is a function of z defined for any point $z \neq z_0$ of the set E. The limit of (7.1) as $z \to z_0$, $z \in E$, provided it exists, is called the *derivative* of the function $f(z)$ at the point z_0 relative to the set E, denoted by $f'_E(z_0)$ or simply $f'(z_0)$, and the function $f(z)$ itself is said to be *differentiable* at the point z_0 relative to the set E.

In the special case where E is an interval (finite or infinite) of the real axis, $f(z)$ is a function of the real variable $z = x$ and takes values which are in general complex:

$$f(z) = f(x) = \varphi(x) + i\psi(x).$$

If $\psi(x) \equiv 0$, i.e., if the values of $f(z)$ are real, our definitions of a derivative and of differentiability reduce to the usual definitions given in elementary calculus. However, if $\psi(x) \not\equiv 0$, then, writing

$$\frac{f(x) - f(x_0)}{x - x_0} = \frac{\varphi(x) - \varphi(x_0)}{x - x_0} + i\frac{\psi(x) - \psi(x_0)}{x - x_0},$$

we conclude that $f'(x)$ exists if and only if the derivatives $\varphi'(x)$ and $\psi'(x)$ exist, and that

$$f'(x) = \varphi'(x) + i\psi'(x)$$

(see Sec. 12). For example, if

$$f(x) = a \cos x + ib \sin x,$$

then

$$f'(x) = -a \sin x + ib \cos x.$$

Introducing the notation $\Delta z = z - z_0$ for the increment of the independent variable and $\Delta_E f(z) = f(z) - f(z_0)$ for the increment of the function $f(z)$ (relative to the set E), we have the following necessary and sufficient condition for differentiability:

THEOREM 7.1. *The function $f(z)$ is differentiable at the point $z_0 \in E$ (relative to the set E) if and only if $f(z)$ can be written in the form*

$$\Delta_E f(z) = A\,\Delta z + \varepsilon(z, z_0)\,\Delta z, \tag{7.2}$$

where $\varepsilon(z, z_0) \to 0$ as $\Delta z \to 0$ ($z \in E$), and A is a constant independent of Δz and ε.

Proof. If $f(z)$ has a derivative $f'_E(z_0)$ at z_0, then, by definition

$$\frac{\Delta_E f(z)}{\Delta z} = f'_E(z_0) + \varepsilon(z, z_0), \tag{7.3}$$

where $\varepsilon(z, z_0) \to 0$ as $\Delta z \to 0$. Multiplying (7.3) by Δz, we find that $\Delta_E f(z)$ can be written in the form (7.2) with $A = f'_E(z_0)$. Conversely, if $\Delta_E f(z)$ can be written in the form (7.2), then dividing by Δz and taking the limit as $\Delta z \to 0$, we find that $f'_E(z_0)$ exists and equals A.

Remark. It is an immediate consequence of (7.2) that if $f(z)$ is differentiable at the point $z_0 \in E$, it is also continuous at z_0.[1]

By the *differential* of a function $f(z)$ which is differentiable at the point $z_0 \in E$, we mean the principal linear part of the increment $\Delta_E f(z)$, i.e., the quantity

$$d_E f(z_0) = f'_E(z_0)\,\Delta z = f'_E(z_0)\,dz, \tag{7.4}$$

where, as in elementary calculus, we set $dz = \Delta z$ (see Problem 7.1). We can then write

$$f'_E(z_0) = \frac{d_E f(z_0)}{dz}. \tag{7.5}$$

For simplicity, we usually omit the subscript E in expressions like (7.2) through (7.5), unless we want to emphasize the role of the set E.

---

[1] For brevity, we shall often omit the phrase "relative to E," which is tacitly assumed in any context like this.

Example. To clarify the role of the set E relative to which the derivative is taken, let E be the real axis and consider the function

$$f(z) = f(x) = x. \tag{7.6}$$

Obviously, the derivative $f'_E(z)$ exists for all $z = x \in E$, since

$$\frac{f(z) - f(z_0)}{z - z_0} = \frac{x - x_0}{x - x_0} = 1 \qquad (z \in E).$$

We now extend the function $f(z)$ to the whole complex plane $\mathcal{C}$, by writing (7.6) for all $z = x + iy \in \mathcal{C}$. Then $f(z)$ is continuous for any $z \in \mathcal{C}$ and coincides with the original function when $z \in E$. However, the difference quotient is now

$$\frac{f(z) - f(z_0)}{z - z_0} = \frac{x - x_0}{(x - x_0) + i(y - y_0)}, \tag{7.7}$$

which has no limit as $z \to z_0 \in \mathcal{C}$. In fact, (7.7) approaches 0 if $z \to z_0$ along the line $x = x_0$, whereas (7.7) approaches 1 if $z \to z_0$ along the line $y = y_0$. In other words, $f'_{\mathcal{C}}(z_0)$ does not exist for any $z \in \mathcal{C}$. Thus, it is already apparent from this simple example that the requirement that $f(z)$ be differentiable regarded as a function of a complex variable $z = x + iy$ is much more stringent than the requirement that $f(z)$ be differentiable only for real values of z.

29. Rules for Differentiating Functions of a Complex Variable

It follows from our definition of a derivative and the properties of limits (see Sec. 12) that the basic differentiation rules familiar from elementary calculus can be extended to the case of functions of a complex variable. We now list some of these rules. In each case, $f(z)$, $f_1(z)$, $f_2(z)$, etc., is assumed to be differentiable at a given point $z \in E$ (relative to E).

Rule 1. If $f(z) \equiv z$, then

$$\frac{df(z)}{dz} \equiv 1,$$

whereas if $f(z) \equiv$ const, then

$$\frac{df(z)}{dz} \equiv 0.$$

Loosely speaking, the derivative of the independent variable is 1, and the derivative of any constant is 0.

Rule 2 (Differentiation of sums, products and powers). If c is a constant, then

$$\frac{d[cf(z)]}{dz} = c\frac{df(z)}{dz}.$$

Moreover, we have

$$\frac{d}{dz}[f_1(z) + f_2(z) + \cdots + f_n(z)] = \frac{df_1(z)}{dz} + \frac{df_2(z)}{dz} + \cdots + \frac{df_n(z)}{dz},$$

$$\frac{d}{dz}[f_1(z)f_2(z)\cdots f_n(z)] = \frac{df_1(z)}{dz}f_2(z)f_3(z)\cdots f_n(z) + f_1(z)\frac{df_2(z)}{dz}f_3(z)\cdots f_n(z) + \cdots + f_1(z)f_2(z)\cdots f_{n-1}(z)\frac{df_n(z)}{dz},$$

$$\frac{d}{dz}[f(z)]^n = n[f(z)]^{n-1}f'(z).$$

In particular,

$$\frac{d}{dz}z^n = nz^{n-1}$$

and

$$\frac{d}{dz}(a_0 + a_1z + a_2z^2 + \cdots + a_nz^n) = a_1 + 2a_2z + \cdots + na_nz^{n-1}.$$

Rule 3 (*Differentiation of a quotient*). If $f_2(z) \neq 0$, then

$$\frac{d}{dz}\left[\frac{f_1(z)}{f_2(z)}\right] = \frac{f_2(z)\dfrac{df_1(z)}{dz} - f_1(z)\dfrac{df_2(z)}{dz}}{[f_2(z)]^2}.$$

Rule 4 (*Differentiation of a composite function*). Suppose the function $f(z)$ with domain E and range $\mathscr{E}$ is differentiable at the point $z_0 \in E$, and suppose $\mathscr{E}$ is an infinite set with $w_0 = f(z_0)$ as a limit point. Moreover, suppose the function $\varphi(w)$ has domain $\mathscr{E}$ and is differentiable at the point $w_0 \in \mathscr{E}$. Then the *composite function*

$$\varphi[f(z)]$$

is differentiable at the point z_0, and

$$\frac{d_E\varphi[f(z_0)]}{dz} = \frac{d_{\mathscr{E}}\varphi(w_0)}{dw}\frac{d_Ef(z_0)}{dz}. \tag{7.8}$$

The proof is trivial if there exists a neighborhood $\mathscr{N}(z_0)$ of the point z_0 such that $f(z) = w \neq w_0$ for all $z \in \mathscr{N}(z_0) \cap E$, for then

$$\frac{\varphi[f(z)] - \varphi[f(z_0)]}{z - z_0} = \frac{\varphi(w) - \varphi(w_0)}{w - w_0}\frac{w - w_0}{z - z_0} = \frac{\varphi(w) - \varphi(w_0)}{w - w_0}\frac{f(z) - f(z_0)}{z - z_0}. \tag{7.9}$$

Since $w \to w_0$ as $z \to z_0$ (see the remark on p. 106), taking the limit of (7.9) as $z \to z_0$, we immediately obtain (7.8).

Now suppose every neighborhood of z_0 contains a point $z \in E$, $z \neq z_0$ such that $f(z) = f(z_0) = w_0$. Then there exists a sequence $\{z_n\}$ of distinct points in E converging to z_0 such that

$$f(z_n) = f(z_0) \qquad (n = 1, 2, \ldots).$$

For this sequence

$$\frac{f(z_n) - f(z_0) = 0}{z_n - z_0} \qquad (n = 1, 2, \ldots),$$

and hence the derivative

$$\frac{d_E f(z)}{dz} = \lim_{z \to \infty} \frac{f(z) - f(z_0)}{z - z_0}$$

(which by hypothesis exists) must vanish. Therefore, the right-hand side of (7.8) vanishes, and to prove (7.8) we have to show that its left-hand side also vanishes. Since the difference quotient

$$\frac{\varphi[f(z)] - \varphi[f(z_0)]}{z - z_0}$$

is obviously zero for every $z \neq z_0$ such that $f(z) = f(z_0)$, the proof reduces to showing that

$$\lim_{n \to \infty} \frac{\varphi[f(z_n^*)] - \varphi[f(z_0)]}{z_n^* - z_0} \tag{7.10}$$

vanishes for every sequence $\{z_n^*\}$ converging to z_0 such that $f(z_n^*) = w_n^* \neq w_0 = f(z_0)$. However, for such a sequence, (7.10) equals

$$\frac{\varphi(w_n^*) - \varphi(w_0)}{w_n^* - w_0} \frac{w_n^* - w_0}{z_n^* - z_0} = \frac{\varphi(w_n^*) - \varphi(w_0)}{w_n^* - w_0} \frac{f(z_n^*) - f(z_0)}{z_n^* - z_0}$$

[cf. (7.9)], which approaches

$$\frac{d_{\mathcal{E}}\varphi(w_0)}{dw} \frac{d_E f(z_0)}{dz}$$

as $n \to \infty$. But as just shown, $d_E f(z_0)/dz$ vanishes, and hence (7.8) holds.

Rule 5 (*Differentiation of an inverse function*). Suppose $w = f(z)$ is a one-to-one function defined on E, and suppose the inverse function $z = f^{-1}(w) = \varphi(w)$ is continuous on $\mathcal{E}$, the range of $f(z)$. Then, if $f(z)$ is differentiable at the point $z_0 \in E$ and if $f'_E(z_0) \neq 0$, the function $\varphi(w)$ is differentiable at the point $w_0 = f(z_0) \in \mathcal{E}$, and

$$\varphi'_{\mathcal{E}}(w_0) = \frac{1}{f'_E(z_0)}.$$

The proof goes as follows: Since the function $w = f(z)$ is one-to-one, $w \neq w_0$ implies $z \neq z_0$, and hence the difference quotient of the function $\varphi(w)$ can be written in the form

$$\frac{\varphi(w) - \varphi(w_0)}{w - w_0} = \frac{z - z_0}{w - w_0} = \frac{1}{\dfrac{w - w_0}{z - z_0}}.$$

Moreover, since $\varphi(w)$ is continuous on $\mathscr{E}$, $\varphi(w) \to \varphi(w_0)$ as $w \to w_0$, i.e., $z \to z_0$ as $w \to w_0$. Therefore

$$\varphi'_{\mathscr{E}}(w_0) = \lim_{w \to w_0} \frac{\varphi(w) - \varphi(w_0)}{w - w_0} = \frac{1}{\lim\limits_{z \to z_0} \dfrac{w - w_0}{z - z_0}} = \frac{1}{\lim\limits_{z \to z_0} \dfrac{f(z) - f(z_0)}{z - z_0}} = \frac{1}{f'_E(z_0)},$$

as asserted.

30. The Cauchy-Riemann Equations. Analytic Functions

From now on, we shall be concerned mainly with functions defined on some domain $E = G$, and we shall drop the subscript E in $f'_E(z)$ and $d_E f(z)/dz$. Suppose $u(x, y)$ is a real function of two real variables x and y defined on G. Then $u(x, y)$ is said to be *differentiable* at the point $(x_0, y_0) \in G$ if $u(x, y)$ can be written in the form[2]

$$\begin{aligned} u(x, y) - u(x_0, y_0) &= A(x_0, y_0)(x - x_0) + B(x_0, y_0)(y - y_0) \\ &\quad + \varepsilon_1(x, y; x_0, y_0)(x - x_0) + \varepsilon_2(x, y; x_0, y_0)(y - y_0), \end{aligned} \tag{7.11}$$

where

$$\varepsilon_1(x, y; x_0, y_0) \to 0, \qquad \varepsilon_2(x, y; x_0, y_0) \to 0$$

as $(x, y) \to (x_0, y_0)$. The coefficients $A(x_0, y_0)$ and $B(x_0, y_0)$ in the right-hand side of (7.11) are just the partial derivatives of the function $u(x, y)$ at the point (x_0, y_0):

$$A(x_0, y_0) = \frac{\partial u(x, y)}{\partial x}\bigg|_{\substack{x = x_0 \\ y = y_0}}, \qquad B(x_0, y_0) = \frac{\partial u(x, y)}{\partial y}\bigg|_{\substack{x = x_0 \\ y = y_0}}.$$

THEOREM 7.2 (*Cauchy-Riemann equations*). *Let*

$$f(z) = u(x, y) + iv(x, y)$$

be a function of a complex variable defined on a domain G. Then a necessary and sufficient condition for $f(z)$ to be differentiable (as a function of a complex variable) at the point $z_0 = x_0 + iy_0 \in G$ is that the functions

[2] See e.g., D. V. Widder, *op. cit.*, p. 17.

$u(x, y)$ and $v(x, y)$ be differentiable (as functions of the two real variables x and y) at the point (x_0, y_0) and satisfy the Cauchy-Riemann equations[3]

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{7.12}$$

at (x_0, y_0). If these conditions are satisfied, $f'(z_0)$ can be represented in any of the forms

$$f'(z_0) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y} = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y} = \frac{\partial v}{\partial y} + i\frac{\partial v}{\partial x}, \tag{7.13}$$

where the partial derivatives are evaluated at (x_0, y_0).

Proof. First we prove that the conditions are necessary. Consider the increments

$$\begin{aligned} \Delta z = z - z_0 &= (x - x_0) + i(y - y_0) = \Delta x + i\,\Delta y, \\ \Delta f(z) = f(z) - f(z_0) &= [u(x, y) - u(x_0, y_0)] + i[v(x, y) - v(x_0, y_0)] \\ &= \Delta u + i\,\Delta v \end{aligned}$$

of the independent variable z and of the function $f(z)$. According to Theorem 7.1, if $f(z)$ is differentiable at z_0,

$$\Delta f(z) = f'(z_0)\,\Delta z + \varepsilon\,\Delta z, \tag{7.14}$$

where $\varepsilon \to 0$ as $\Delta z \to 0$. Writing

$$f'(z_0) = a + ib, \qquad \varepsilon = \varepsilon_1 + i\varepsilon_2,$$

and taking the real and imaginary parts of (7.14), we find that

$$\begin{aligned} \Delta u &= a\,\Delta x - b\,\Delta y + \varepsilon_1\,\Delta x - \varepsilon_2\,\Delta y, \\ \Delta v &= b\,\Delta x + a\,\Delta y + \varepsilon_2\,\Delta x + \varepsilon_1\,\Delta y, \end{aligned}$$

where $\varepsilon_1, \varepsilon_2 \to 0$ as $\Delta x, \Delta y \to 0$, since

$$|\Delta z| = \sqrt{(\Delta x)^2 + (\Delta y)^2}, \quad |\varepsilon_1| \leqslant |\varepsilon|, \quad |\varepsilon_2| \leqslant |\varepsilon|.$$

It follows that the functions $u(x, y)$ and $v(x, y)$ are differentiable at (x_0, y_0) and

$$\frac{\partial u}{\partial x} = a, \quad \frac{\partial u}{\partial y} = -b, \quad \frac{\partial v}{\partial x} = b, \quad \frac{\partial v}{\partial y} = a. \tag{7.15}$$

But (7.15) immediately implies (7.12) and (7.13).

---

[3] The equations (7.12), which are usually called the *Cauchy-Riemann equations*, are of basic importance in the theory of analytic functions and its applications to problems of physics and engineering. However, it should be noted that this universally encountered attribution is not historically justified. In fact, the equations (7.12) had already been studied in the eighteenth century by D'Alembert (1717–1783) and Euler (1707–1783), in researches devoted to the application of functions of a complex variable to hydrodynamics (D'Alembert and Euler), and to cartography and integral calculus (Euler).

To prove that the conditions of the theorem are sufficient, we reverse the preceding argument. Thus, suppose $u(x, y)$ and $v(x, y)$ are differentiable at the point (x_0, y_0), and suppose (7.12) holds. Then

$$\Delta u = \frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y + \alpha_1 \Delta x + \alpha_2 \Delta y = a\,\Delta x - b\,\Delta y + \alpha_1 \Delta x + \alpha_2 \Delta y,$$
$$\Delta v = \frac{\partial v}{\partial x}\Delta x + \frac{\partial v}{\partial y}\Delta y + \beta_1 \Delta x + \beta_2 \Delta y = b\,\Delta x + a\,\Delta y + \beta_1 \Delta x + \beta_2 \Delta y, \tag{7.16}$$

where $\alpha_1, \alpha_2, \beta_1, \beta_2 \to 0$ as $\Delta x, \Delta y \to 0$ and we have written

$$a = \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad b = -\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x}. \tag{7.17}$$

Substituting (7.16) into the formula $\Delta f(z) = \Delta u + i\,\Delta v$, we find that

$$\begin{aligned}\Delta f(z) &= a(\Delta x + i\,\Delta y) + ib(\Delta x + i\,\Delta y) + (\alpha_1 + i\beta_1)\Delta x + (\alpha_2 + i\beta_2)\Delta y \\ &= (a + ib)\,\Delta z + \left[(\alpha_1 + i\beta_1)\frac{\Delta x}{\Delta z} + (\alpha_2 + i\beta_2)\frac{\Delta y}{\Delta z}\right]\Delta z\end{aligned}$$

or

$$\Delta f(z) = (a + ib)\,\Delta z + \varepsilon\,\Delta z, \tag{7.18}$$

where

$$\begin{aligned}|\varepsilon| &= \left|(\alpha_1 + i\beta_1)\frac{\Delta x}{\Delta z} + (\alpha_2 + i\beta_2)\frac{\Delta y}{\Delta z}\right| \\ &\leqslant |\alpha_1 + i\beta_1|\left|\frac{\Delta x}{\Delta z}\right| + |\alpha_2 + i\beta_2|\left|\frac{\Delta y}{\Delta z}\right| \\ &\leqslant |\alpha_1 + i\beta_1| + |\alpha_2 + i\beta_2| \leqslant |\alpha_1| + |\beta_1| + |\alpha_2| + |\beta_2|.\end{aligned}$$

Since $\alpha_1, \alpha_2, \beta_1, \beta_2 \to 0$ as $\Delta x, \Delta y \to 0$, it follows that $\varepsilon \to 0$ as $\Delta z \to 0$, and hence, according to (7.18) and Theorem 7.1, $f(z)$ is differentiable at z_0, with derivative

$$f'(z_0) = a + ib. \tag{7.19}$$

The various representations (7.13) of $f'(z_0)$ are now immediate consequences of (7.19) and (7.17).

A function $f(z)$ which is differentiable on a domain G, i.e., at every point of G, is said to be *analytic* (synonymously, *holomorphic* or *regular*) *on* G. If $f(z)$ is analytic in a neighborhood of z_0, $f(z)$ is said to be *analytic at* z_0. The term *holomorphic* was introduced by Cauchy's students Briot (1817–1882) and Bouquet (1819–1895).[4] The meaning of the term *analytic*, used earlier by Lagrange (1736–1813), later by Weierstrass (1815–1897) and commonly accepted today, was explained in Sec. 1.

[4] The word *holomorphic*, stemming from the Greek ὅλος = *entire* and μορφή = *form*, was intended to stress the resemblance between holomorphic functions and *entire functions*, i.e., functions which (like polynomials) are differentiable everywhere in the finite plane.

The use of this term to describe functions of a complex variable which are differentiable on a domain G will be justified in Sec. 81, where it is shown that such a function can be represented as the sum of a convergent power series in some neighborhood of an arbitrary point of G. For the time being, however, we shall use the phrase *analytic function* as a synonym for a function of a complex variable which is differentiable on a given domain (or in a neighborhood of a given point).

Remark. As we know from calculus,[5] a sufficient condition for the differentiability of the functions $u(x, y)$ and $v(x, y)$ on a domain G is that the partial derivatives

$$\frac{\partial u}{\partial x}, \quad \frac{\partial u}{\partial y}, \quad \frac{\partial v}{\partial x}, \quad \frac{\partial v}{\partial y} \tag{7.20}$$

exist and be continuous on G. Therefore, a sufficient condition for the function $f(z) = u + iv$ to be analytic on G is that the partial derivatives (7.20) exist, be continuous and satisfy the equations (7.12) on G.

Example 1. For the function

$$f(z) = e^x(\cos y + i \sin y), \tag{7.21}$$

defined on the whole plane, we have

$$u = e^x \cos y, \qquad v = e^x \sin y,$$

with continuous partial derivatives

$$\frac{\partial u}{\partial x} = e^x \cos y = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -e^x \sin y = -\frac{\partial v}{\partial x}.$$

Therefore the Cauchy-Riemann equations (7.12) are satisfied, and the function (7.21) is analytic in the whole plane, with derivative

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = e^x \cos y + ie^x \sin y = f(z).$$

Example 2. For the function $f(z) = x$, considered in the example on p. 107, we have

$$u = x, \qquad v = 0$$

and

$$\frac{\partial u}{\partial x} = 1, \quad \frac{\partial u}{\partial y} = 0, \quad \frac{\partial v}{\partial x} = 0, \quad \frac{\partial v}{\partial y} = 0.$$

Since

$$\frac{\partial u}{\partial x} \neq \frac{\partial v}{\partial y},$$

the Cauchy-Riemann equations are not satisfied, and hence this function is not differentiable anywhere in the plane.

---

[5] See e.g., D. V. Widder, *op. cit.*, p. 17.

In many cases, it is important to express the differentiability conditions for a function $f(z) = u + iv$ at a point $z \neq 0$ in terms of the polar coordinates

$$r = |z|, \qquad \Phi = \operatorname{Arg} z.$$

The appropriate necessary and sufficient conditions for differentiability are that $u(r, \Phi)$ and $v(r, \Phi)$ be differentiable (as functions of the two real variables r and Φ) and satisfy the polar form of the Cauchy-Riemann equations, i.e,

$$\frac{\partial u}{\partial r} = \frac{1}{r}\frac{\partial v}{\partial \Phi}, \qquad \frac{\partial v}{\partial r} = -\frac{1}{r}\frac{\partial u}{\partial \Phi} \tag{7.22}$$

(at a given nonzero point P). To verify these conditions, we must prove that 1) u and v are differentiable as functions of r and Φ at P if and only if they are differentiable as functions of x and y at P, and 2) under these conditions, the equations (7.22) are equivalent to the equations (7.12). The first assertion follows from the familiar fact that a differentiable function ($u = u(x, y)$, say) of differentiable functions ($x = r\cos\Phi, y = r\sin\Phi$, say) is also differentiable (with respect to the new variables r and Φ).[6] The second assertion can be verified by direct calculation. For example, if u and v are differentiable functions of x and y, and if the equations (7.12) hold, then

$$\begin{aligned}
\frac{\partial u}{\partial r} &= \frac{\partial u}{\partial x}\cos\Phi + \frac{\partial u}{\partial y}\sin\Phi = \frac{\partial v}{\partial y}\cos\Phi - \frac{\partial v}{\partial x}\sin\Phi = \frac{1}{r}\frac{\partial v}{\partial \Phi}, \\
\frac{\partial v}{\partial r} &= \frac{\partial v}{\partial x}\cos\Phi + \frac{\partial v}{\partial y}\sin\Phi = -\frac{\partial u}{\partial y}\cos\Phi + \frac{\partial u}{\partial x}\sin\Phi = -\frac{1}{r}\frac{\partial u}{\partial \Phi}.
\end{aligned} \tag{7.23}$$

Writing (7.23) in the form

$$\begin{aligned}
\frac{\partial u}{\partial r} &= \frac{\partial u}{\partial x}\cos\Phi - \frac{\partial v}{\partial x}\sin\Phi, \\
\frac{\partial v}{\partial r} &= \frac{\partial u}{\partial x}\sin\Phi + \frac{\partial v}{\partial x}\cos\Phi,
\end{aligned}$$

and solving for $\partial u/\partial x$ and $\partial u/\partial y$, we obtain

$$\begin{aligned}
\frac{\partial u}{\partial x} &= \frac{\partial u}{\partial r}\cos\Phi + \frac{\partial v}{\partial r}\sin\Phi, \\
\frac{\partial v}{\partial x} &= -\frac{\partial u}{\partial r}\sin\Phi + \frac{\partial v}{\partial r}\cos\Phi,
\end{aligned}$$

and hence

$$\begin{aligned}
f'(z) &= \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial u}{\partial r}(\cos\Phi - i\sin\Phi) + i\frac{\partial v}{\partial r}(\cos\Phi - i\sin\Phi) \\
&= \left(\frac{\partial u}{\partial r} + i\frac{\partial v}{\partial r}\right)(\cos\Phi - i\sin\Phi) = \frac{r}{z}\left(\frac{\partial u}{\partial r} + i\frac{\partial v}{\partial r}\right).
\end{aligned} \tag{7.24}$$

---

[6] See e.g., R. C. Buck, *op. cit.*, Theorem 14, p. 191.

This formula is convenient for calculating $f'(z)$ with the help of polar coordinates. Using (7.22), we can write $f'(z)$ in the form

$$f'(z) = \frac{1}{z}\left(\frac{\partial v}{\partial \Phi} - i\frac{\partial u}{\partial \Phi}\right).$$

Example. Consider the function

$$\begin{aligned} f(z) = z^{m/n} &= |z|^{m/n}\left(\cos\frac{m\ \mathrm{Arg}\ z}{n} + i\sin\frac{m\ \mathrm{Arg}\ z}{n}\right) \\ &= r^{m/n}\left(\cos\frac{m\Phi}{n} + i\sin\frac{m\Phi}{n}\right), \end{aligned} \tag{7.25}$$

where m and $n > 0$ are integers (see Sec. 6). This function is defined on the domain G consisting of all nonzero points of the complex plane, and is

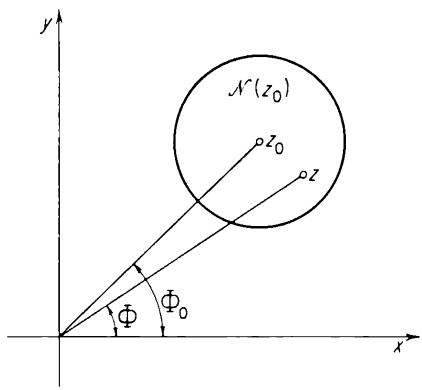

FIGURE 7.1

multiple-valued (unless m/n is an integer), since $\Phi = \mathrm{Arg}\ z$ is multiple-valued. Before we can talk about the derivative of $z^{m/n}$, we must first make $z^{m/n}$ single-valued in the following sense: Let $z_0 \in G$, and choose a neighborhood $\mathcal{N}(z_0)$ which does not contain the origin. Fix one of the values of $\Phi_0 = \mathrm{Arg}\ z_0$, and for the argument of any other point $z \in \mathcal{N}(z_0)$ choose the unique value Φ satisfying the condition

$$|\Phi - \Phi_0| < \frac{\pi}{2}$$

(see Figure 7.1).[7] Using this value of Φ in (7.25), we obtain a single-valued function defined on $\mathcal{N}(z_0)$, which we call a *single-valued* (and *continuous*)

[7] A similar construction was made in Sec. 9 (cf. Figure 3.3).

branch of the function (7.25), denoted by $z^{m/n}$ as before.[8] Thus at any point $z \in \mathcal{N}(z_0)$, including z_0 itself, we can now write

$$u = r^{m/n} \cos \frac{m\Phi}{n}, \qquad v = r^{m/n} \sin \frac{m\Phi}{n}$$

and

$$\frac{\partial u}{\partial r} = \frac{m}{n} r^{(m/n)-1} \cos \frac{m\Phi}{n} = \frac{1}{r} \frac{\partial v}{\partial \Phi},$$
$$\frac{\partial v}{\partial r} = \frac{m}{n} r^{(m/n)-1} \sin \frac{m\Phi}{n} = -\frac{1}{r} \frac{\partial u}{\partial \Phi},$$

i.e., $f(z)$ satisfies the equations (7.22), and is therefore differentiable on $\mathcal{N}(z_0)$. According to formula (7.24),

$$f'(z) = \frac{r}{z}\left(\frac{m}{n} r^{(m/n)-1} \cos \frac{m\Phi}{n} + i \frac{m}{n} r^{(m/n)-1} \sin \frac{m\Phi}{n}\right)$$
$$= \frac{m}{n} r^{m/n}\left(\cos \frac{m\Phi}{n} + i \sin \frac{m\Phi}{n}\right) \frac{1}{z} = \frac{m}{n} \frac{f(z)}{z},$$

so that the rule for differentiating a fractional power $z^{m/n}$ is formally the same as the rule for differentiating the corresponding function $x^{m/n}$ of a real variable. It should be kept in mind that our calculation is subject to the condition $z \neq 0$, which can only be dropped if m/n is a nonnegative integer.

PROBLEMS

7.1. Justify the formula $dz = \Delta z$ ("the differential and the increment of the independent variable are equal") and hence the formula (7.4).

Hint. Cf. Rule 1, p. 107.

7.2. Show that the function $f(z) = \bar{z}$ is not differentiable anywhere in the z-plane.

7.3. Show that the function $f(z) = z \operatorname{Re} z$ is differentiable only at the point $z = 0$, and find $f'(0)$.

Comment. Thus $f(z) = z \operatorname{Re} z$ is differentiable but not analytic at $z = 0$.

7.4. Show that the function

$$u(x, y) = \begin{cases} x & \text{if } |y| > |x|, \\ -x & \text{if } |y| \leqslant |x| \end{cases}$$

has partial derivatives $\partial u/\partial x$ and $\partial u/\partial y$ at the origin, but is not differentiable there.

7.5. Show that the function $f(z) = \sqrt{|xy|}$ satisfies the Cauchy-Riemann equations at the point $z = 0$, but is not differentiable there.

---

[8] For further details on single-valued branches of multiple-valued functions, see Chap. 11.

7.6. Establish the following generalization of the Cauchy-Riemann equations: If $f(z) = u + iv$ is differentiable at a point $z_0 = x_0 + iy_0$ of a domain G, then

$$\frac{\partial u}{\partial s} = \frac{\partial v}{\partial n}, \qquad \frac{\partial u}{\partial n} = -\frac{\partial v}{\partial s}$$

at (x_0, y_0), where $\partial/\partial s$ and $\partial/\partial n$ denote directional differentiation in any two orthogonal directions s and n at (x_0, y_0) such that n is obtained from s by making a counterclockwise rotation.

7.7. Let the function $f(z) = u + iv$ be such that both u and v are differentiable at a point (x_0, y_0), and suppose that

$$\lim_{z \to z_0} \left| \frac{f(z) - f(z_0)}{z - z_0} \right|$$

exists at $z_0 = x_0 + iy_0$. Prove that either $f(z)$ or $\overline{f(z)}$ is differentiable at z_0.

7.8. Let the function $f(z) = u + iv$ be such that both u and v are differentiable at a point (x_0, y_0), and suppose that

$$\lim_{z \to z_0} \arg \frac{f(z) - f(z_0)}{z - z_0}$$

exists at $z_0 = x_0 + iy_0$. Prove that $f(z)$ is differentiable at z_0.

7.9. Prove that if $f(z)$ is analytic on a domain G and if $f'(z) \equiv 0$ on G, then $f(z) \equiv$ const on G.

Hint. Use the fact that any two points of G can be joined by a polygonal curve contained in G (Theorem 4.12), and then apply Theorem 3.14 to show that $f(z)$ takes the same value at the end points of any line segment contained in G.

7.10. Show that the equations (7.22) imply the equations (7.12) if u and v are differentiable functions of r and Φ.

7.11. Deduce the equations (7.22) from the generalized Cauchy-Riemann equations of Problem 7.6.

7.12. Let G be the domain consisting of all nonzero points of the finite complex plane. Prove that the function $f(z) = \ln r + i\Phi$, where $r = |z|$ and $\Phi = \operatorname{Arg} z$, is analytic on G, with derivative $1/z$.

Hint. First choose an appropriate single-valued continuous branch of $f(z)$ [cf. Sec. 56].

8

GEOMETRIC INTERPRETATION OF THE DERIVATIVE. CONFORMAL MAPPING

31. Geometric Interpretation of $\operatorname{Arg} f'(z)$

Let l be a continuous curve with equation $z = \lambda(t)$, $t \in [a, b]$, and suppose $\lambda(t)$ is differentiable at a point $t_0 \in [a, b]$ (relative to the set $[a, b]$). Let $\{t_n\}$ be an arbitrary sequence of points in $[a, b]$ converging to t_0 ($t_n \neq t_0$), and consider the difference quotient

$$r_n = \frac{\lambda(t_n) - \lambda(t_0)}{t_n - t_0}. \tag{8.1}$$

Obviously $r_n \to r_0 = \lambda'(t_0)$ as $n \to \infty$.

DEFINITION. *The curve l is said to have a tangent at the point $z_0 = \lambda(t_0)$ if the limit*

$$0 = \lim_{n \to \infty} \operatorname{Arg} r_n \tag{8.2}$$

exists, and then the tangent is said to have inclination 0. Geometrically, the tangent to l at z_0 is represented by the ray τ emanating from z_0 which makes the angle 0 with the positive real axis.[1]

[1] Formula (8.2) means that given any $\varepsilon > 0$, there is an integer $N(\varepsilon) > 0$ and a sequence $\{0_n\}$, where each 0_n is a value of $\operatorname{Arg} r_n$, such that $|0_n - 0| < \varepsilon$ for all $n > N(\varepsilon)$. Clearly, 0 is only defined to within a multiple of 2π. The angle 0 will always be measured from the positive real axis to the tangent τ (in the counterclockwise direction for a positive value of 0).

Remark 1. Clearly, if $\lambda'(t_0) \neq 0$, l has a tangent at z_0, since

$$\theta = \lim_{n \to \infty} \operatorname{Arg} r_n = \operatorname{Arg} r_0 = \operatorname{Arg} \lambda'(t_0)$$

(see p. 35), and then the inclination of the tangent is just the argument of the complex number $\lambda'(t_0)$. On the other hand, if $\lambda'(t_0) = 0$, l may or may not have a tangent at z_0, since the fact that $r_n \to 0$ implies only that $|r_n| \to 0$ and says nothing about the behavior of $\operatorname{Arg} r_n$ (see p. 33). However, if l has a tangent at z_0, then $\lambda(t_n) \neq \lambda(t_0)$ for all t_n sufficiently close to t_0, since Arg 0 is meaningless.[2]

Remark 2. As we have defined it, the tangent is a ray, not a vector. If $\lambda'(t_0) \neq 0$, we can also introduce a *tangent vector* to l at z_0, defined as the vector of length $|\lambda'(t_0)|$ which makes the angle θ with the positive real axis.

THEOREM 8.1. *Let G be a domain, and let $f(z)$ be a continuous function of a complex variable defined on G. Suppose $f(z)$ has a nonzero derivative $f'(z_0)$ at a point $z_0 \in G$, and let l be a curve which passes through z_0 and has a tangent τ at z_0. Then $w = f(z)$ maps l into a curve L in the w-plane which passes through the point $w_0 = f(z_0)$ and has a tangent T at w_0. Moreover, the inclination of T exceeds the inclination of τ by the angle* $\operatorname{Arg} f'(z_0)$.

Proof. Suppose l has the equation $z = \lambda(t)$, $t \in [a, b]$, and let $z_0 = \lambda(t_0)$. By hypothesis,

$$\theta = \lim_{n \to \infty} \operatorname{Arg} r_n$$

exists, where r_n is given by (8.1). The function $w = f(z)$ maps l into a curve L in the w-plane with equation

$$w = f[\lambda(t)] = \Lambda(t), \qquad t \in [a, b], \tag{8.3}$$

where $w_0 = f(z_0) = \Lambda(t_0)$. Let $\{t_n\}$ be an arbitrary sequence of points in $[a, b]$ converging to t_0, and let

$$R_n = \frac{\Lambda(t_n) - \Lambda(t_0)}{t_n - t_0}.$$

Then the tangent to L at w_0 has inclination

$$\Theta = \lim_{n \to \infty} \operatorname{Arg} R_n,$$

provided this limit exists. Clearly we have

$$R_n = \frac{\Lambda(t_n) - \Lambda(t_0)}{\lambda(t_n) - \lambda(t_0)} \frac{\lambda(t_n) - \lambda(t_0)}{t_n - t_0},$$

[2] This condition is automatically satisfied if $\lambda'(t_0) \neq 0$, since otherwise $r_n = 0$ for t arbitrarily close to t_0, which implies that $\lambda'(t_0) = 0$, contrary to hypothesis.

where the first factor in the right-hand side is well defined, since $\lambda(t) \neq \lambda(t_0)$ for all t_n sufficiently close to t_0 (l is assumed to have a tangent at z_0). Therefore

$$\begin{aligned}\Theta &= \lim_{n\to\infty} \operatorname{Arg} R_n = \lim_{n\to\infty} \operatorname{Arg} \frac{\Lambda(t_n) - \Lambda(t_0)}{\lambda(t_n) - \lambda(t_0)} \frac{\lambda(t_n) - \lambda(t_0)}{t_n - t_0} \\ &= \lim_{n\to\infty} \operatorname{Arg} \frac{w_n - w_0}{z_n - z_0} r_n = \lim_{n\to\infty} \operatorname{Arg} \frac{w_n - w_0}{z_n - z_0} + \lim_{n\to\infty} \operatorname{Arg} r_n \qquad (8.4) \\ &= \operatorname{Arg} \lim_{n\to\infty} \frac{w_n - w_0}{z_n - z_0} + \theta = \operatorname{Arg} f'(z_0) + \theta,\end{aligned}$$

where $w_n = \Lambda(t_n)$, $z_n = \lambda(t_n)$ and θ is the inclination of τ at z_0.[3] It follows from (8.4) that Θ exists and that

$$\Theta - \theta = \operatorname{Arg} f'(z_0),$$

as asserted. We note that things are particularly simple in the case where $\lambda'(t_0) \neq 0$, since then

$$\begin{aligned}\Theta &= \operatorname{Arg} \Lambda'(t_0) = \operatorname{Arg} [f'(z_0)\lambda'(t_0)] \\ &= \operatorname{Arg} f'(z_0) + \operatorname{Arg} \lambda'(t_0) = \operatorname{Arg} f'(z_0) + \theta,\end{aligned}$$

by the rule for differentiating the composite function (8.3).

Now let l_1 and l_2 be two curves with a common initial point z_0, which have tangents τ_1 and τ_2 at z_0, and suppose the angle between τ_1 and τ_2 is measured from τ_1 to τ_2. Suppose l_1 and l_2 have images L_1 and L_2 under $f(z)$. Then, according to Theorem 8.1, if $f'(z_0) \neq 0$, L_1 and L_2 have tangents T_1 and T_2 at the point $w_0 = f(z_0)$, where T_1 and T_2 are obtained by rotating τ_1 and τ_2 through the same angle $\operatorname{Arg} f'(z_0)$. Therefore the angle between L_1 and L_2 equals the angle between l_1 and l_2, and is measured in the same direction, i.e., from L_1 to L_2. In other words, a continuous function $w = f(z)$ with a nonzero derivative $f'(z_0)$ maps all curves in the z-plane which pass through z_0 and have tangents at z_0 into curves in the w-plane which pass through $w_0 = f(z_0)$ and have tangents at w_0, and moreover, *the mapping preserves angles between curves.* A mapping by a continuous function which preserves angles between curves passing through a given point z_0 is said to be *conformal at* z_0 (cf. p. 88). If a conformal mapping preserves the directions in which angles are measured (as well as their magnitudes), it is called a *conformal mapping of the first kind,* but if it reverses the directions in which angles are measured, it is called a *conformal mapping of the second kind.* If a mapping is conformal at all points of a domain G, it is said to be *conformal on* G. Thus Theorem 8.1 has the following consequence:

---

[3] In reversing the order of the operations Arg and $\lim_{n\to\infty}$, we have used the fact that $f'(z_0) \neq 0$.

THEOREM 8.2. *Let G be a domain, and let $f(z)$ be an analytic function on G. Then $f(z)$ is a conformal mapping of the first kind at every point of G where $f'(z) \neq 0$.*

Example 1. Reflection in the real axis, i.e., the transformation $w = \bar{z}$, is a conformal mapping of the second kind. A more general example is the complex conjugate

$$w = \overline{f(z)}$$

of an analytic function $f(z)$, where $f'(z) \neq 0$.

Example 2. At a point where the derivative vanishes, angles may or may not be preserved, as can be seen by comparing the mappings

$$f_1(z) = r^2(\cos \Phi + i \sin \Phi) = rz,$$
$$f_2(z) = r^2(\cos 2\Phi + i \sin 2\Phi) = z^2$$

at the point $z = 0$.

32. Geometric Interpretation of $|f'(z)|$

As we have just seen, $\operatorname{Arg} f'(z_0)$ represents the rotation undergone by the tangent to a curve l at the point $z_0 \in l$ when transforming to the new curve $L = f(l)$ and the new point $w_0 = f(z_0)$. In particular, if $f'(z_0)$ is a positive real number, the tangents to l at z_0 and to L at w_0 are parallel and point in the same direction.

To explain the geometric meaning of the quantity $|f'(z_0)|$, i.e., the absolute value of the derivative at z_0, we note that

$$|f'(z_0)| = \lim_{z \to z_0} \frac{|f(z) - f(z_0)|}{|z - z_0|}.$$

The numbers $|z - z_0|$ and $|f(z) - f(z_0)|$ are the distance between the points z and z_0 in the z-plane, and the distance between their images $f(z)$ and $f(z_0)$ in the w-plane, respectively. Thus, interpreting

$$\frac{|f(z) - f(z_0)|}{|z - z_0|}$$

as the *linear magnification ratio* (or simply the *magnification*) of the vector $z - z_0$ under the mapping $w = f(z)$,[4] we can regard $|f'(z_0)|$ as the *magnification at the point* z_0 under $w = f(z)$.

---

[4] Here the word *magnification* is used in a general sense, and can correspond to *stretching* if $|f'(z_0)| > 1$ or *shrinking* if $|f'(z_0)| < 1$ [or neither if $|f'(z_0)| = 1$].

Remark. The size of the magnification at the point z_0 does not depend on the choice of the finite vector $z - z_0$ drawn from z_0, since $|f'(z_0)|$ is not the actual magnification of any such vector, but rather the limiting magnification as $z \to z_0$.

33. The Mapping $w = \dfrac{az + b}{cz + d}$

To illustrate the above considerations, we now examine the *fractional linear transformation* or *Möbius transformation*

$$L(z) = \frac{az + b}{cz + d}, \tag{8.5}$$

where a, b, c, d are arbitrary complex numbers (except that c and d are not both zero). First suppose that $c = 0$. Then $L(z)$ reduces to

$$L(z) = \alpha z + \beta \qquad (\alpha = a/d, \quad \beta = b/d), \tag{8.6}$$

and is sometimes called the *entire linear transformation.* The transformation (8.6) is defined for all values of z, and if $\alpha \neq 0$, its derivative $L'(z)$ is a nonzero constant, so that (8.6) is a conformal mapping of the whole z-plane. Under this transformation, the tangents to all curves in the z-plane are rotated through the same angle, equal to Arg α, and the magnification at every point equals $|\alpha|$. If $\alpha = 1$, then

$$\operatorname{Arg} \alpha = 2k\pi, \qquad |\alpha| = 1,$$

where k is an integer, and then both the rotation and expansion produce no effect. In this case, the transformation takes the form

$$w = z + \beta,$$

which obviously corresponds to displacing the whole plane by the vector β. On the other hand, if $\alpha \neq 1$ (and $\alpha \neq 0$), the transformation (8.6) can be written in the form

$$w - z_0 = \alpha(z - z_0),$$

where z_0 is determined from the equation[5]

$$z_0 = \alpha z_0 + \beta.$$

Then it is immediately clear that the transformation (8.6) is equivalent to a rotation of the whole plane through the angle Arg α about the point

[5] Obviously, z_0 is invariant under the transformation (8.6), i.e., z_0 is a *fixed point* of the transformation. If $\alpha \neq 0$, the point at infinity is also a fixed point (see Sec. 46).

$z_0 = \beta/(1 - \alpha)$, together with a uniform magnification by the factor $|\alpha|$ relative to the point z_0. This magnification is sometimes called a *homothetic transformation* (or *transformation of similitude*) with *ray center* z_0 and *ray ratio* $|\alpha|$.

Next suppose that $c \neq 0$ in (8.4). Then the derivative

$$L'(z) = \frac{ad - bc}{(cz + d)^2} = \frac{ad - bc}{c^2} \frac{1}{(z - \delta)^2}$$

exists, if $z \neq \delta$, where $\delta = -d/c$. If the determinant

$$ad - bc = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

vanishes, then its rows are proportional,[6] i.e.,

$$\frac{a}{c} = \frac{b}{d} = \mu \quad \text{or} \quad a = \mu c, \quad b = \mu d,$$

where μ is a constant, so that (8.5) reduces to the trivial transformation

$$L(z) = \frac{az + b}{cz + d} = \frac{\mu cz + \mu d}{cz + d} \equiv \mu.$$

If $ad - bc \neq 0$, then $L'(z) \neq 0$ for all $z \neq \delta$, and hence the mapping $w = L(z)$ is conformal at all finite points except possibly at $z = \delta$. Under the mapping, the tangents to curves passing through any point $z \neq \delta$ are rotated through an angle equal to

$$\operatorname{Arg} L'(z) = \operatorname{Arg} \frac{ad - bc}{c^2} - 2 \operatorname{Arg} (z - \delta),$$

while the magnification at z equals

$$|L'(z)| = \frac{ad - bc}{c^2} \frac{1}{|z - \delta|^2}.$$

The angle through which tangents are rotated has the same value for all points with equal values of $\operatorname{Arg} (z - \delta)$, i.e., along any ray drawn from δ, but otherwise varies from point to point. Similarly, in general the magnification varies with z, but it has the same value for all points with equal values of $|z - \delta|$, i.e., along any circle with center δ. In particular, the magnification is equal to 1 at every point of the circle C with equation

$$|z - \delta| = \frac{1}{|c|} \sqrt{|ad - bc|}$$

---

[6] See e.g., G. E. Shilov, *op. cit.*, p. 25.

(called the *isometric circle* of the Möbius transformation), is greater than 1 inside C (approaching ∞ as $z \to \delta$), and is less than 1 outside C (approaching 0 as $z \to \infty$). The situation is shown schematically in Figure 8.1.

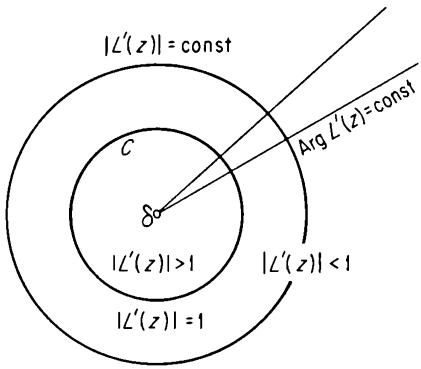

FIGURE 8.1

34. Conformal Mapping of the Extended Plane

As in the preceding section, let $c \neq 0$ and $ad - bc \neq 0$. Then it is clear that

$$\lim_{z \to \delta} \frac{az + b}{cz + d} = \infty, \qquad \lim_{z \to \infty} \frac{az + b}{cz + d} = \frac{a}{c} = A,$$

and hence

$$L(\delta) = \infty, \qquad L(\infty) = A.$$

Thus the function $w = L(z)$ maps the finite point δ into the point at infinity (∞), and maps the point at infinity into the finite point A. We now show that the mapping is conformal at these points too. First let γ_1 and γ_2 be two curves forming an angle θ with its vertex at the point δ, and let Γ_1 and Γ_2 be their images in the w-plane. To prove that Γ_1 and Γ_2 form an angle θ with its vertex at infinity, we subject the w-plane to the transformation

$$\eta = \frac{1}{w}.$$

Then the curves Γ_1 and Γ_2 go into two curves Γ_1^* and Γ_2^*, and the point at infinity goes into the origin of coordinates (see Figure 8.2). Obviously, we can go from γ_1 and γ_2 in the z-plane to Γ_1^* and Γ_2^* in the η-plane by making the Möbius transformation

$$\eta = \frac{1}{w} = \frac{cz + d}{az + b},$$

which is conformal at the point $z = \delta = -d/c$. It follows that the curves Γ_1^* and Γ_2^* form an angle θ with its vertex at the origin. Therefore, according

to the definition given in Sec. 25, the curves Γ_1 and Γ_2 also form an angle θ with its vertex at infinity. This proves that the mapping $w = L(z)$ is conformal at the point $z = \delta$.

The fact that $w = L(z)$ is conformal at ∞ is proved similarly. In fact, if the curves γ_1 and γ_2 go through the point at infinity in the z-plane, their images Γ_1 and Γ_2 in the w-plane go through the point A. Suppose γ_1 and γ_2 form an angle θ with its vertex at infinity. This means that their images γ_1^*

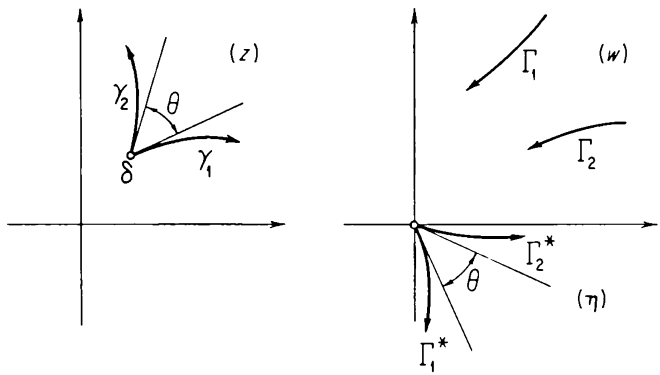

FIGURE 8.2

and γ_2^* under the transformation $\zeta = 1/z$ form an angle θ with its vertex at the origin. But we can obviously go from γ_1^* and γ_2^* to Γ_1 and Γ_2 by making the Möbius transformation

$$w = \frac{az + b}{cz + d} = \frac{a\dfrac{1}{\zeta} + b}{c\dfrac{1}{\zeta} + d} = \frac{a + b\zeta}{c + d\zeta},$$

which is conformal at the point $\zeta = 0$. It follows that Γ_1 and Γ_2 form the same angle θ at the point $A = a/c$. This proves that the mapping $w = L(z)$ is conformal at ∞. The situation can be summarized by saying that the *transformation $w = L(z)$ is a conformal mapping of the extended plane onto itself.*

Remark 1. These considerations suggest the following definition: A function $f(z)$ is said to be *analytic at $z = \infty$* if the function $f^*(\zeta) = f(1/\zeta)$ is analytic at $\zeta = 0$. In particular, if $f(z)$ is analytic at $z = \infty$, the limit

$$\lim_{z \to \infty} f(z) = \lim_{\zeta \to 0} f^*(\zeta) = f(\infty)$$

always exists and is finite. We define the derivative of $f(z)$ at $z = \infty$ to be the quantity

$$f'(\infty) = f^{*\prime}(0),$$

where it should be noted that in general

$$f'(\infty) \neq \lim_{z \to \infty} f'(z)$$

(see Problem 8.10). Then, by the argument given above for the special case of the Möbius transformation, it is easy to see that the mapping $w = f(z)$ is conformal at ∞ if $f'(\infty) \neq 0$. With this approach, the conformality at ∞ of the Möbius transformation

$$L(z) = \frac{az + b}{cz + d} \qquad (ad - bc \neq 0, \quad c \neq 0)$$

follows from the fact that

$$L'(\infty) = -\frac{ad - bc}{c^2} \neq 0.$$

Remark 2. Similarly, if

$$\lim_{z \to a} f(z) = \infty,$$

but if

$$\varphi(z) = \frac{1}{f(z)}$$

is analytic at $z = a$, with derivative $\varphi'(a) \neq 0$, then, just as in the case of the Möbius transformation, the mapping $w = f(z)$ is conformal at $z = a$.

PROBLEMS

8.1. With the same notation as on p. 118, a curve l is said to have a *left-hand tangent* (of inclination θ) at the point $z_0 = \lambda(t_0)$ if the limit (8.2) exists, subject to the extra condition that every point of the sequence $\{t_n\}$ converging to t_0 be less than t_0. The *right-hand tangent* is defined similarly by requiring that $t_n > t_0$ for every n. Give an example of a (continuous) curve l which has a left-hand tangent but no right-hand tangent (and hence no tangent) at a point $z_0 \in l$.

8.2. Verify that the function

$$f_1(z) = r^2(\cos \Phi + i \sin \Phi) = rz$$

used in Example 2, p. 121 is differentiable at $z = 0$.

8.3. Find the angle through which a curve drawn from the point z_0 is rotated under the mapping $w = z^2$ if

a) $z_0 = i$; b) $z_0 = -\frac{1}{4}$; c) $z_0 = 1 + i$; d) $z_0 = -3 + 4i$.

Also find the corresponding values of the magnification.

8.4. Carry out the same calculations as in the preceding problem, this time applied to the function $w = z^3$.

8.5. Which part of the plane is shrunk and which part stretched under the following mappings:

a) $w = z^2$; b) $w = z^2 + 2z$; c) $w = \frac{1}{z}$?

8.6. As shown on pp. 122–123, the entire linear transformation $w = \alpha z + \beta$ is equivalent to a rotation and a magnification relative to the fixed point $z_0 = \beta/(1 - \alpha)$, provided that $\alpha \neq 0$. Find the rotation, magnification and (finite) fixed point, if such exists, corresponding to each of the following transformations, and write each in the canonical form $w - z_0 = \alpha(z - z_0)$:

a) $w = 2z + 1 - 3i$; b) $w = iz + 4$;
c) $w = z + 1 - 2i$; d) $w - w_1 = a(z - z_1)$ $(a \neq 0)$.

8.7. Find the entire linear transformation with fixed point $1 + 2i$ carrying the point i into the point $-i$.

Ans. $w = (2 + i)z + 1 - 3i$.

8.8. Find the entire linear transformation carrying the triangle with vertices at the points $0, 1, i$ into the similar triangle with vertices at the points $0, 2, 1 + i$.

8.9. Prove that the transformation $L(z) = \alpha z + \beta$ is conformal at infinity if $\alpha \neq 0$.

8.10. Prove that if $f(z)$ is analytic at infinity, then

$$\lim_{z \to \infty} f'(z) = 0.$$

9

ELEMENTARY ENTIRE FUNCTIONS

35. Polynomials

The simplest class of differentiable functions of a complex variable is the class of single-valued functions which are analytic everywhere in the finite plane.[1] Such functions are known as *entire functions*. An important (but very special) subclass of the class of entire functions consists of *polynomials*, i.e., functions of the form

$$P_n(z) = a_0 + a_1 z + \cdots + a_n z^n \qquad (a_n \neq 0), \tag{9.1}$$

where n is a nonnegative integer, called the *degree* of $P_n(z)$, and $a_0, a_1, \ldots, a_n$ are finite complex numbers, called the *coefficients* of $P_n(z)$. If $n = 0$, $P_n(z)$ reduces to a constant, but if $n > 0$,

$$\lim_{n \to \infty} P_n(z) = \infty, \qquad \text{i.e.,} \quad P(\infty) = \infty.$$

According to the *fundamental theorem of algebra* (see p. 8, Theorem 17.7 and Problem 17.14), the equation

$$P_n(z) = 0 \qquad (n > 0)$$

always has a complex root, say α_1. It follows that

$$P_n(z) = (z - \alpha_1) P_{n-1}(z),$$

[1] It will be shown later (see Theorem 16.8 and Problem 16.12) that any function which is analytic everywhere in the *extended plane* must be the trivial function $f(z) \equiv$ const.

where $P_{n-1}(z)$ is a polynomial of degree $n-1$ (see Problems 9.1 and 9.2). If $P_{n-1}(\alpha_1) = 0$, then by the same argument,

$$P_{n-1}(z) = (z - \alpha_1)P_{n-2}(z),$$

where $P_{n-2}(z)$ is a polynomial of degree $n-2$, and hence

$$P_n(z) = (z - \alpha_1)^2 P_{n-2}(z).$$

Continuing this process k_1 times ($1 \leqslant k_1 \leqslant n$), we eventually find that

$$P_n(z) = (z - \alpha_1)^{k_1} P_{n-k_1}(z),$$

where $P_{n-k_1}(z)$ is a polynomial of degree $n - k_1$ and $P_{n-k_1}(\alpha_1) \neq 0$. The number α_1 is called a *root of multiplicity* k_1 (or a k_1*-fold root*) of the equation $P_n(z) = 0$; it is also called a *zero* (or *zero-point*) *of order* k_1 of the polynomial $P_n(z)$.[2] If $k_1 < n$, then, applying the fundamental theorem of algebra again, we find that $P_{n-k_1}(z)$ has a zero, say α_2, where $\alpha_2 \neq \alpha_1$. Therefore, by the previous argument,

$$P_n(z) = (z - \alpha_1)^{k_1}(z - \alpha_2)^{k_2} P_{n-k_1-k_2}(z),$$

where $1 \leqslant k_2 \leqslant n - 1$ and $P_{n-k_1-k_2}(\alpha_2) \neq 0$. Repeating this process as many times as necessary (but no more than n times), we finally arrive at the unique factorization

$$P_n(z) = a_n(z - \alpha_1)^{k_1} \cdots (z - \alpha_p)^{k_p}, \tag{9.2}$$

where $1 \leqslant p \leqslant n$ and $k_1 + \cdots + k_p = n$ (see Problem 9.3).

Now let A be an arbitrary finite complex number. Then, since $P_n(z) - A$ is obviously a polynomial of degree n, the equation

$$P_n(z) = A$$

also has n solutions (which are not necessarily distinct). As for the equation

$$P_n(z) = \infty, \tag{9.3}$$

it has only the solution $z = \infty$, since $P_n(z) = \infty$ if and only if $z = \infty$. For consistency, we formally regard $z = \infty$ as an n-fold root of the equation (9.3).[3]

It follows that *every polynomial* $w = P_n(z)$ *of degree* $n > 0$ *maps the extended plane onto itself in such a way that every point* w *has at most* n *distinct inverse images.* Moreover, only certain exceptional values of w (always including ∞ if $n > 1$) can have fewer than n distinct inverse images:

THEOREM 9.1. *If* $P_n(z)$ *is a polynomial of degree* $n > 1$, *there are at most* n *points in the extended* w*-plane with fewer than* n *distinct inverse images under the mapping* $w = P_n(z)$.

---

[2] This is the preferred terminology, but one can also talk of the *order* of a root or the *multiplicity* of a zero.

[3] This is also in keeping with the fact that $\zeta = 0$ is a pole of order n of the rational function $P_n(1/\zeta)$ [see Sec. 43].

Proof. Since $n > 1$, the point ∞ has fewer than n inverse images (in fact, just one). If $A \neq \infty$ has fewer than n distinct inverse images, the equation

$$P_n(z) = A$$

must have a multiple root. As is well known (see Problem 9.5) any multiple root of this equation satisfies the equation

$$P_n'(z) = 0,$$

where $P_n'(z)$ is the derivative of $P_n(z)$. But this equation, being of degree $n - 1$, has no more than $n - 1$ distinct roots $\gamma_1, \ldots, \gamma_r$ $(1 \leqslant r \leqslant n - 1)$. Therefore the numbers

$$P_n(\gamma_1), \ldots, P_n(\gamma_r), \infty \qquad (1 \leqslant r \leqslant n - 1)$$

are the only values of A for which the equation $P_n(z) = A$ can have a multiple root. Since at most n of these numbers are distinct, the theorem is proved. (If $n = 0$ or $n = 1$, the theorem is meaningless.)

36. The Mapping $w = P_n(z)$

According to the general theory of Chap. 8, the mapping $w = P_n(z)$ is conformal at all points of the z-plane, with the possible exception of the point at infinity and the points $\gamma_1, \ldots, \gamma_r$ $(1 \leqslant r \leqslant n - 1)$ at which the derivative $P_n'(z)$ vanishes. In the case $n = 1$, $w = P_n(z)$ reduces to the entire linear transformation (8.5), and the mapping is one-to-one and conformal everywhere in the extended plane (cf. Problem 8.10). However, for $n > 1$ the mapping actually fails to be conformal at the points $\gamma_1, \ldots, \gamma_2, \infty$, as we now show:

THEOREM 9.2. *Let z_0 be a root of multiplicity $k > 1$ of the equation $P_n(z) = P_n(z_0)$. Then, under the mapping $w = P_n(z)$, every angle with its vertex at z_0 is enlarged k times. Moreover, every angle with its vertex at infinity is enlarged n times.*

Proof. If z_0 is a root of multiplicity $k > 1$ of the equation

$$P_n(z) = P_n(z_0),$$

and hence a root of multiplicity $k - 1$ of the equation

$$P_n'(z) = 0,$$

then

$$P_n(z) - P_n(z_0) = (z - z_0)^k Q(z), \tag{9.4}$$

where $Q(z)$ is a polynomial which does not vanish for $z = z_0$. It follows from (9.4) and the properties of the function Arg z (see Sec. 5) that

$$\operatorname{Arg} \frac{w - w_0}{(z - z_0)^k} = \operatorname{Arg} Q(z), \tag{9.5}$$

where $w = P_n(z)$, $w_0 = P_n(z_0)$. Now let l be a curve in the z-plane, with equation $z = \lambda(t)$, $t \in [a, b]$, which has a tangent τ with inclination θ at the point $z_0 = \lambda(t_0)$, $t_0 \in [a, b]$. Then the image of l under $f(z)$ is a curve L in the w-plane, with equation

$$w = P_n[\lambda(t)] = \Lambda(t), \qquad t \in [a, b].$$

Let $\{t_n\}$ be an arbitrary sequence of points in $[a, b]$ converging to t_0 ($t_n \neq t_0$). By a familiar argument (see Sec. 31), the inclination of the tangent to L at w_0 (if such exists) equals

$$\Theta = \lim_{n \to \infty} \operatorname{Arg} \frac{\Lambda(t_n) - \Lambda(t_0)}{t_n - t_0} = \lim_{n \to \infty} \operatorname{Arg} \frac{w_n - w_0}{(z_n - z_0)^k} \frac{(z_n - z_0)^k}{t_n - t_0}, \tag{9.6}$$

where $w_n = \Lambda(t_n)$, $w_0 = \Lambda(t_0)$, $z_n = \lambda(t_n)$, $z_0 = \lambda(t_0)$. Moreover $z_n \to z_0$ as $n \to \infty$, but $z_n \neq z_0$ for t_n sufficiently close to t_0 (since τ exists); this last fact has been anticipated in writing (9.6). Since $Q(z_0) \neq 0$, it follows from (9.6), (9.5) and the definition of θ that Θ exists and equals

$$\Theta = \lim_{n \to \infty} \operatorname{Arg} Q(z_n) + k \lim_{n \to \infty} \frac{\lambda(t_n) - \lambda(t_0)}{t_n - t_0} = \operatorname{Arg} Q(z_0) + k\theta \tag{9.7}$$

(cf. the proof of Theorem 8.1).[4]

Next let l_1 and l_2 be two curves in the z-plane with a common initial point z_0, which have tangents τ_1 and τ_2 at z_0. Then the angle with vertex z_0 formed by l_1 and l_2 (measured from l_1 to l_2) equals

$$\delta = \theta_2 - \theta_1,$$

where θ_1 and θ_2 are the inclinations of τ_1 and τ_2, respectively. According to (9.7), the images L_1 and L_2 of l_1 and l_2 under the mapping $w = P_n(z)$ form an angle

$$\Delta = \Theta_2 - \Theta_1 = [\operatorname{Arg} Q(z_0) + k\theta_2] - [\operatorname{Arg} Q(z_0) + k\theta_1] = k(\theta_2 - \theta_1) = k\delta,$$

with its vertex at the point $w_0 = P_n(z_0)$. This proves the first part of the theorem.

To prove the second part of the theorem, we make use of the transformation $\zeta = 1/z$, obtaining the new function

$$\eta = P_n^*(\zeta) = \frac{1}{P_n(1/\zeta)} = \frac{\zeta^n}{a_n + a_{n-1}\zeta + \cdots + a_0\zeta^n} \qquad (a_n \neq 0).$$

Then the point $\zeta = 0$ corresponds to the point $z = \infty$, and we have to prove that under the mapping $\eta = P_n^*(\zeta)$, every angle in the ζ-plane with

[4] Note that even if $\lambda'(t_0) \neq 0$, we still cannot write $\Theta = \operatorname{Arg} \Lambda'(t_0)$, since $\Lambda'(t_0) = P_n'(z_0)\lambda'(t_0) = 0$.

its vertex at the origin is enlarged n times. Let l be a curve in the ζ-plane, with equation $\zeta = \lambda(t)$, $t \in [a, b]$, which goes through the point $\zeta = 0$ and has a tangent τ with inclination 0 at $\zeta = 0$. Then, by a slight modification of the previous argument, we find that the curve L in the η-plane with equation $\eta = P_n^*[\lambda(t)] = \Lambda(t)$ goes through the point $\eta = 0$ and has a tangent with inclination

$$\Theta = n0 - \operatorname{Arg} a_n$$

at $\eta = 0$. The rest of the proof follows as before (note that $w = 1/\eta$).

37. The Mapping $w = (z - a)^n$

We now make a detailed study of the mapping

$$w = (z - a)^n \qquad (n > 1). \tag{9.8}$$

This function maps the extended z-plane onto the extended w-plane in such a way that every point w has n distinct inverse images, with the exception of the two points $w = 0$ and $w = \infty$, for which the n inverse images "coalesce" into the single points $z = a$ and $z = \infty$, respectively. To find the n inverse images of w when $w \neq 0$, $w \neq \infty$, we solve (9.8) for z, obtaining

$$z = a + \sqrt[n]{w} = a + \sqrt[n]{|w|}\left(\cos\frac{\operatorname{Arg} w}{n} + i\sin\frac{\operatorname{Arg} w}{n}\right). \tag{9.9}$$

Obviously, the n distinct points (9.9) lie at the vertices of a regular n-gon with its center at the point $z = a$. The mapping (9.8) is conformal at all points except $z = a$, $z = \infty$, and every angle with its vertex at one of these two points is enlarged n times.

To get a clearer picture of the mapping (9.8), we observe that

$$|w| = |z - a|^n, \qquad \operatorname{Arg} w = n \operatorname{Arg}(z - a),$$

which implies that every circle of radius r with its center at the point $z = a$ is mapped into a circle of radius r^n with its center at the point $w = 0$. Moreover, as the point z goes around the circle $|z - a| = r$ once in the positive direction [so that $\operatorname{Arg}(z - a)$ increases continuously by 2π], the image point w goes around the circle $|w| = r^n$ n times in the same direction [since $\operatorname{Arg} w$ increases continuously by $2n\pi$]. We also note that as the point z sweeps out the ray

$$\operatorname{Arg}(z - z) = \varphi_0 + 2k\pi \qquad (k = 0, \pm 1, \pm 2, \ldots)$$

going from a to ∞, the image point w sweeps out the ray

$$\operatorname{Arg} w = n\varphi_0 + 2m\pi \qquad (m = 0, \pm 1, \pm 2, \ldots)$$

going from 0 to ∞.

Next consider the domain G consisting of all points z such that

$$\varphi_0 + 2k\pi < \operatorname{Arg}(z - a) < \varphi_1 + 2k\pi \qquad (k = 0, \pm 1, \pm 2, \ldots),$$

where $0 < \varphi_1 - \varphi_0 \leqslant 2\pi/n$. Such a domain will be called the *interior* of the angle of $\varphi_1 - \varphi_0$ radians formed by the rays

$$\operatorname{Arg}(z - a) = \varphi_0 + 2k\pi, \quad \operatorname{Arg}(z - a) = \varphi_1 + 2k\pi \quad (k = 0, \pm 1, \pm 2, \ldots),$$

the term *angle* itself being reserved either for the figure formed by these two rays (an "unbounded curve") or for the quantity $\varphi_1 - \varphi_0$. Then the image of G under the mapping (9.8) is the domain

$$n\varphi_0 + 2m\pi < \operatorname{Arg} w < n\varphi_1 + 2m\pi \qquad (m = 0, \pm 1, \pm 2, \ldots),$$

i.e., the interior of the angle of $n(\varphi_1 - \varphi_0)$ radians with its vertex at the origin of the w-plane (see Figure 9.1). Not only is the function $w = (z - a)^n$

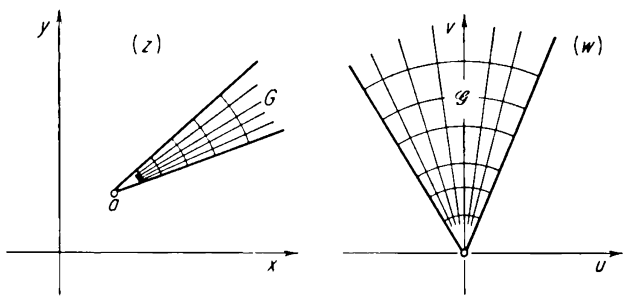

FIGURE 9.1

conformal on G, as already noted, but it is also one-to-one on G. In fact, since $w = (z - a)^n$ is single-valued, we need only verify that every point w has only one inverse image in G. Since the n inverse images of the point w lie at the vertices of a regular n-gon in the z-plane with center at a, two inverse images can belong to the interior of the same angle with vertex at a only if the angle exceeds $2\pi/n$. But $0 < \varphi_1 - \varphi_0 \leqslant 2\pi/n$ by hypothesis, and hence every point of $\mathscr{G}$ has only one inverse image in G, as asserted. Thus, *the function $w = (z - a)^n$ is a one-to-one conformal mapping of the interior of one angle onto the interior of another angle which is n times larger.*

Of course, it would be quite incorrect to conclude that the function $w = (z - a)^n$ maps every straight line into a straight line, and every circle into a circle. For example, suppose $a = 0$, $n = 2$, so that $w = (z - a)^n$ reduces to $w = z^2$, and consider the effect of this mapping on straight lines parallel to the coordinate axes. Every line parallel to the imaginary axis (except the axis itself) has an equation of the form

$$z = b + it \qquad (-\infty < t < \infty),$$

where $b \neq 0$ is a real number, and the image of this line under the mapping $w = z^2$ has the equation

$$w = (b + it)^2. \tag{9.10}$$

Separating real and imaginary parts of (9.10), we obtain the corresponding parametric equations

$$u = b^2 - t^2, \qquad v = 2bt,$$

where u and v are the rectangular coordinates of the point w. Elimination of the parameter t leads to

$$v^2 = 4b^2(b^2 - u), \tag{9.11}$$

which is the equation of a parabola opening to the left, with axis lying along the real axis and focus at the origin. Similarly, every line parallel to the real axis (except the axis itself) has an equation of the form

$$z = t + ic \qquad (-\infty < t < \infty),$$

where $c \neq 0$ is a real number, and it is easily verified that the image of this line under the mapping $w = z^2$ is the parabola

$$v^2 = 4c^2(c^2 + u) \tag{9.12}$$

opening to the right, with axis lying along the real axis and focus at the origin. Thus, *the mapping $w = z^2$ transforms the two one-parameter families of straight lines parallel to (but distinct from) the coordinate axes into two one-parameter families of parabolas* (9.11) *and* (9.12), *with axes lying along the real axis and a common focus at the origin* (see Figure 9.2). Moreover, since the two families of straight lines in the z-plane form an *orthogonal system* (i.e., every curve in one family is orthogonal to every curve in the other family, and *vice versa*), and since the mapping $w = z^2$ is conformal (except at $z = 0$), it follows that the two families of parabolas in the w-plane also form an orthogonal system.[5]

Remark. It should be kept in mind that $w = z^2$ is not a one-to-one mapping, and in fact, every point in the w-plane except $w = 0$ and $w = \infty$ has two inverse images. In particular, the inverse image of the parabola (9.11) consists of the two straight lines $z = b + it$ and $z = -b + it$ which are symmetric with respect to the imaginary axis, while the inverse image of the parabola (9.12) consists of the two straight lines $z = t + ic$ and $z = t - ic$ which are symmetric with respect to the real axis. However, if we confine ourselves to a half-plane G whose boundary consists of a line passing through the origin (i.e., the interior of an angle of π radians with vertex at the origin),

[5] The x and y-axes themselves are orthogonal, but their images, i.e., the line segments $u \geqslant 0$, $v = 0$ and $u \leqslant 0$, $v = 0$ (each traversed twice) are not orthogonal, since the derivative of $w = z^2$ vanishes at $z = 0$ (cf. Theorem 9.2).

then the correspondence between G and its image $\mathcal{G}$ under $w = z^2$ is one-to-one. In fact, $\mathcal{G}$ is the interior of an angle of 2π radians with vertex at the origin, both of whose sides coalesce to form a single ray emanating from the origin.

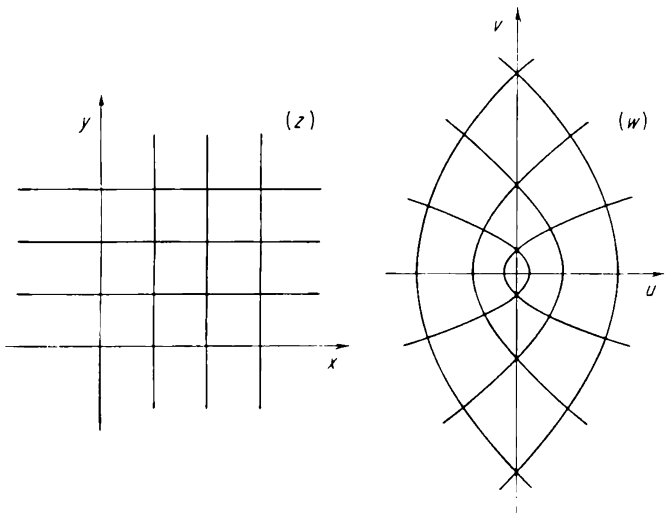

FIGURE 9.2

38. The Exponential

An entire function which is not a polynomial is called an *entire transcendental function*. The simplest example of such a function is the *exponential (function)* e^z or $\exp z$, obtained by suitably extending the familiar function e^x, defined for the real variable x, to the case where x takes arbitrary complex values, i.e., is replaced by the complex variable $z = x + iy$. It is not hard to show that the real exponential $f(x) = e^x$ is the unique function with the following properties:

1. $f(x)$ is defined and single-valued for all real x, takes only real values, and in particular takes the value e when $x = 1$;
2. $f(x)$ satisfies the *addition theorem*

$$f(x_1 + x_2) = f(x_1)f(x_2);$$

3. $f(x)$ is continuous for all x.[6]

The complex function $f(z) = e^z$ can be characterized in much the same way:

[6] This result follows at once by a specialization of the proof of Theorem 9.3.

THEOREM 9.3. *There is a unique function* $f(z) = e^z$ *with the following properties*:

1'. $f(z)$ *is defined and single-valued for all (finite) complex* z, *takes real values when* z *is real, and in particular takes the value* e *when* $z = 1$;

2'. $f(z)$ *satisfies the addition theorem*

$$f(z_1 + z_2) = f(z_1)f(z_2); \tag{9.13}$$

3'. $f(z)$ *is differentiable for all* z, *i.e.*, $f(z)$ *is entire.*[7]

Proof. Since

$$f(z)f(1 - z) = f(1) = e \neq 0,$$

the function $f(z)$ cannot vanish for any z. Therefore $|f(z)| > 0$ and $\operatorname{Arg} f(z)$ are defined for all z (the latter function is defined only to within an arbitrary integral multiple of 2π). It follows from (9.13) that

$$|f(z_1 + z_2)| = |f(z_1)|\,|f(z_2)|$$

and

$$\operatorname{Arg} f(z_1 + z_2) = \operatorname{Arg} f(z_1) + \operatorname{Arg} f(z_2),$$

and hence the real functions of a complex variable $\ln |f(z)|$ and $\operatorname{Arg} f(z)$ both satisfy the functional equation

$$F(z_1 + z_2) = F(z_1) + F(z_2). \tag{9.14}$$

When $F(z)$ is the multiple-valued function $\operatorname{Arg} f(z)$, equation (9.14) means that given any value of $F(z_1 + z_2)$, values of $F(z_1)$ and $F(z_2)$ can be found whose sum is $F(z_1 + z_2)$, and conversely, given any values of $F(z_1)$ and $F(z_2)$, their sum is one of the values of $F(z_1 + z_2)$.

It follows from property 3' that $f(z)$ is continuous, and therefore $\ln |f(z)|$ and $\operatorname{Arg} f(z)$ are also continuous. As applied to $\operatorname{Arg} f(z)$, this means that given any value of $\operatorname{Arg} z$ and any sequence $\{z_n\}$ converging to z, there is a sequence of values of $\operatorname{Arg} z_n$ such that $\operatorname{Arg} f(z_n) \to \operatorname{Arg} f(z)$ as $n \to \infty$ (cf. footnote 1, p. 118). Thus $\ln |f(z)|$ and $\operatorname{Arg} f(z)$ are to be found among the continuous real solutions of (9.14). To find these solutions, we need only solve (9.14) for the case where the independent variable is restricted to real values. In fact,

$$F(x + iy) = F(x) + F(iy), \tag{9.15}$$

where $F(x)$ and $F(iy)$ are functions of the real variables x and y, respectively, satisfying the equations

$$F(x_1 + x_2) = F(x_1) + F(x_2),$$
$$F[i(y_1 + y_2)] = F(iy_1) + F(iy_2).$$

---

[7] It will become apparent in the course of the proof that the assumption that $f(z)$ is merely *continuous* for all z is not sufficient to determine $f(z)$ completely (unlike the real case).

Therefore, once we know $F(x)$ and $F(iy)$ separately, we can immediately write down $F(x + iy)$ by forming the sum (9.15).

First suppose $F(x)$ is a single-valued continuous real function, and let $F(1) = a$. Then

$$\underbrace{F\left(\frac{1}{n}\right) + \cdots + F\left(\frac{1}{n}\right)}_{n \text{ times}} = F(1) = a,$$

which implies

$$F\left(\frac{1}{n}\right) = \frac{a}{n}$$

and

$$\underbrace{F\left(\frac{1}{n}\right) + \cdots + F\left(\frac{1}{n}\right)}_{m \text{ times}} = F\left(\frac{m}{n}\right) = \frac{m}{n}a.$$

Moreover

$$F(0) = 0,$$

since

$$F(x + 0) = F(x) + F(0) = F(x).$$

It follows that

$$F(x) + F(-x) = F(0) = 0,$$

and hence

$$F\left(-\frac{m}{n}\right) = -\frac{m}{n}a.$$

Therefore $F(x) = ax$ for all rational x. If $x = \xi$ is irrational, and if $\{x_n\}$ is a sequence of rational numbers converging to ξ, we find that

$$F(\xi) = \lim_{n \to \infty} F(x_n) = a\xi,$$

since $F(x)$ is continuous. Thus we see that $F(x) = ax$ for all real x. Similarly, we have $F(iy) = by$, so that

$$F(z) = F(x) + F(iy) = ax + by.$$

Moreover, since F is real, so are the constants a and b.

Next suppose $F(x)$ is a multiple-valued continuous real function, all of whose values are obtained from one value by adding arbitrary integral multiples of 2π, where $F(x)$ is continuous in the sense just described in connection with the function $\operatorname{Arg} f(z)$. Then from the relation

$$F(x) + F(0) = F(x),$$

we deduce that the values of $F(0)$ are all integral multiples of 2π, and hence that there exists a value of $F(0)$ equal to 0. Therefore, since $F(x)$ is continuous, there must exist a $\delta > 0$ such that

$$|F(x)| < \frac{\pi}{2} \tag{9.16}$$

for all $|x| \leqslant \delta$. Because of the way in which $F(x)$ is multiple-valued, it follows that $F(x)$ has only one value satisfying (9.16). Let this value be denoted by $\Phi(x)$. Obviously, $\Phi(x)$ is continuous (in the ordinary sense) on the interval $[-\delta, \delta]$. If x_1 and x_2 are two numbers in $[-\delta, \delta]$ such that $x_1 + x_2$ also belongs to $[-\delta, \delta]$, then $\Phi(x_1) + \Phi(x_2)$, being one of the values of $F(x_1) + F(x_2) = F(x_1 + x_2)$, can differ from $\Phi(x_1 + x_2)$ only by a multiple of 2π. But

$$|\Phi(x_1) + \Phi(x_2)| < \pi, \qquad |\Phi(x_1 + x_2)| < \pi,$$

and therefore

$$\Phi(x_1 + x_2) = \Phi(x_1) + \Phi(x_2). \tag{9.17}$$

Writing $\Phi(\delta) = \alpha\delta$ and applying our previous argument to the single-valued continuous function $\Phi(x)$ satisfying the functional equation (9.17), we find that

$$\Phi(x) = \alpha x$$

for $|x| \leqslant \delta$. If we use this formula to *define* $\Phi(x)$ for arbitrary real x, then, as before, one of the values of $F(x)$ equals $\Phi(x)$. In fact, any $x \geqslant 0$ can be uniquely represented in the form $2n\delta + \xi$, where n is a nonnegative integer and $|\xi| \leqslant \delta$. Therefore, if $x \geqslant 0$, one value of the sum

$$\underbrace{F(\delta) + \cdots + F(\delta)}_{2n \text{ times}} + F(\xi) = F(x)$$

is $2n\Phi(\delta) + \Phi(\xi) = \Phi(x)$. In other words, if $x \geqslant 0$, all the values of $F(x)$ are of the form

$$\Phi(x) + 2j\pi \qquad (j = 0, \pm 1, \pm 2, \ldots).$$

On the other hand, if $x \leqslant 0$, we deduce from the relation

$$F(x) + F(-x) = F(0) = 2k\pi \qquad (k = 0, \pm 1, \pm 2, \ldots)$$

that

$$\begin{aligned} F(x) = 2k\pi - F(-x) &= 2k\pi - \Phi(-x) - 2j\pi = \Phi(x) + 2(k - j)\pi \\ &= \Phi(x) + 2l\pi \qquad (k, j, l = 0, \pm 1, \pm 2, \ldots). \end{aligned}$$

Thus, if $F(x)$ is a multiple-valued function of the type under discussion, we have

$$F(x) = \alpha x + 2j\pi \qquad (j = 0, \pm 1, \pm 2, \ldots)$$

for all real x. Similarly, we have

$$F(iy) = \beta y + 2k\pi \qquad (k = 0, \pm 1, \pm 2, \ldots),$$

so that

$$F(z) = F(x) + F(iy) = \alpha x + \beta y + 2l\pi \qquad (l = 0, \pm 1, \pm 2, \ldots).$$

Moreover, since F is real, so are the constants α and β.

It follows from these considerations that any single-valued continuous real function satisfying (9.14) must have the form $ax + by$, whereas any multiple-valued continuous real function satisfying (9.14), with the "same kind of multiple-valuedness" as $\operatorname{Arg} f(z)$, must have the form $\alpha z + \beta y + 2l\pi$ ($l = 0, \pm 1, +2, \ldots$), where a, b, α and β are all real constants. Therefore, any continuous complex function which satisfies property 2′ (the addition theorem), and is not identically zero or unity, must be of the form

$$\begin{aligned} f(z) &= e^{\ln|f(z)|}\, e^{i \operatorname{Arg} f(z)} \\ &= e^{ax+by}\,[\cos(\alpha x + \beta y) + i \sin(\alpha x + \beta y)]. \end{aligned} \tag{9.18}$$

We now apply property 1′, which states that $f(z)$ takes real values when z is real. Setting $z = x$, $y = 0$ in (9.18), we see that $\cos \alpha x + i \sin \alpha x$ must be real for all x, which implies $\alpha = 0$. Moreover, property 1′ also requires that $f(1) = e$, so that $e^a = e$ or $a = 1$. Therefore, any continuous function with the properties 1′ and 2′ must have the form

$$f(z) = e^{x+by}(\cos \beta y + i \sin \beta y). \tag{9.19}$$

Finally we impose the full force of property 3′ upon $f(z)$, i.e., we require that $f(z)$ be *differentiable*. Applying the Cauchy-Riemann equations to the function (9.19), we find that

$$\begin{aligned} e^{x+by} \cos \beta y &= be^{x+by} \sin \beta y + \beta e^{x+by} \cos \beta y, \\ be^{x+by} \cos \beta y - \beta e^{x+by} \sin \beta y &= -e^{x+by} \sin \beta y, \end{aligned}$$

or

$$b \sin \beta y + (\beta - 1) \cos \beta y = 0, \tag{9.20}$$

$$b \cos \beta y - (\beta - 1) \sin \beta y = 0. \tag{9.21}$$

Each of these relations by itself is sufficient to determine b and β. In fact, setting first $y = 0$ and then $y = \pi/2$ in (9.20), we find that $\beta = 1$ and $b = 0$, and these values also satisfy (9.21). Thus, finally, the unique function $f(z) = e^z$ with all the properties listed in the statement of the theorem is

$$f(z) \equiv e^z = e^x(\cos y + i \sin y). \tag{9.22}$$

Moreover, the fact that (9.22) has all the desired properties is obvious by inspection.

Remark. If instead of requiring $f(z)$ to be differentiable (property 3′), we had only required $f(z)$ to be continuous, then we would have obtained an infinite set of functions of the form (9.19), where b and β are arbitrary real numbers. However, only the function (9.22) is entire, i.e., analytic everywhere in the finite plane.

39. The Mapping $w = e^z$

It follows from (9.22) that e^z is nonzero for all z and

$$|e^z| = e^x, \qquad \operatorname{Arg} e^z = y + 2k\pi.$$

For $z = iy$ ($x = 0$) we obtain *Euler's formula*

$$e^{iy} = \cos y + i \sin y \tag{9.23}$$

(see p. 8). Using (9.23), we can replace the trigonometric form of a complex number

$$z = r(\cos \Phi + i \sin \Phi)$$

by the more concise *polar form*

$$z = re^{i\Phi}.$$

It is apparent from (9.22) that the exponential is *periodic* in z with *period* $2\pi i$. In other words, if z is changed by $2\pi i$, so that y is changed by 2π, the value of e^z does not change:

$$e^{z+2\pi i} = e^z.$$

We now show that $2\pi i$ is the *fundamental* (or *primitive*) *period* of the function e^z, i.e., that any other period ω of e^z must be of the form $2k\pi i$, where k is an integer. To see this, let $\omega = \alpha + i\beta$. Then

$$e^{z+\omega} = e^z$$

for any z, and in particular,

$$e^{\omega} = e^{\alpha+i\beta} = e^{\alpha}(\cos \beta + i \sin \beta) = 1$$

for $z = 0$. But this means that $|e^{\omega}| = e^{\alpha} = 1$ which implies $\alpha = 0$, and hence $\cos \beta + i \sin \beta = 1$ which implies $\beta = 2k\pi i$, so that

$$\omega = \alpha + i\beta = 2k\pi i,$$

as asserted.

The expression e^{∞} will be regarded as meaningless, since

$$\lim_{z \to \infty} e^z$$

does not exist. This can be seen from the fact that $e^x \to \infty$ as $x \to +\infty$, whereas $e^x \to 0$ as $x \to -\infty$. In particular, it follows that e^z cannot coincide with any polynomial, i.e., e^z is actually an entire *transcendental* function, since any polynomial (excluding the trivial case of a constant) approaches infinity as $z \to \infty$.

Next we study the geometric behavior of the mapping $w = e^z$. As already noted, e^z is nonzero for all z. This means that the origin of coordinates in the w-plane does not belong to the image of the finite z-plane under the mapping $w = e^z$. However, as we now show, any other finite point of the w-plane does belong to this image. In fact, from the equation $w = e^z$, where $w \neq 0$ is given and $z = x + iy$ is unknown, we obtain

$$|w| = e^x \quad \text{or} \quad x = \ln |w|$$

and

$$\operatorname{Arg} w = y + 2k\pi \quad \text{or} \quad y = \operatorname{Arg} w.$$

Therefore the inverse images of the point w can only be points of the form

$$z = \ln |w| + i \operatorname{Arg} w. \tag{9.24}$$

Obviously there are infinitely many points (9.24), since $\operatorname{Arg} w$ takes infinitely many values, all differing by integral multiples of 2π. Moreover, each of these points is actually an inverse image of w, since

$$\begin{aligned} \exp [\ln |w| + i \operatorname{Arg} w] &= e^{\ln |w|} (\cos \operatorname{Arg} w + i \sin \operatorname{Arg} w) \\ &= |w| (\cos \operatorname{Arg} w + i \sin \operatorname{Arg} w) = w. \end{aligned}$$

Therefore the set of all roots of the equation $e^z = w$ $(w \neq 0)$ is given by the formula

$$z = \ln |w| + i \operatorname{Arg} w = \ln |w| + i(\arg w + 2k\pi), \tag{9.25}$$

where $k = 0, \pm 1, \pm 2, \ldots$. These points all lie on the same straight line parallel to the imaginary axis, and the distance between any two consecutive points along the line is 2π. Thus the function $w = e^z$ maps the finite z-plane onto the domain obtained from the finite w-plane by deleting the single point $w = 0$, but the mapping is not one-to-one, since every point $w \neq 0$ has an infinite number of inverse images (9.25). On the other hand, the mapping is conformal at every point of the finite z-plane, since the derivative

$$(e^z)' = \frac{\partial (e^x \cos y)}{\partial x} + i \frac{\partial (e^x \sin y)}{\partial x} = e^x (\cos y + i \sin y) = e^z$$

does not vanish for any value of z.

Now suppose z traces out a straight line parallel to one of the coordinate axes (see Figure 9.3). For example, consider the line

$$z = b + it, \tag{9.26}$$

parallel to the imaginary axis. Then the image of (9.26) under the mapping $w = e^z$ is the curve

$$w = e^b (\cos t + i \sin t), \tag{9.27}$$

i.e., w traces out a circle of radius e^b with its center at the origin. Moreover, as z describes the line (9.26) once in such a way that t, the ordinate of z, increases continuously from $-\infty$ to $+\infty$, w describes the circle (9.27) an infinite number of times in the positive (counterclockwise) direction.

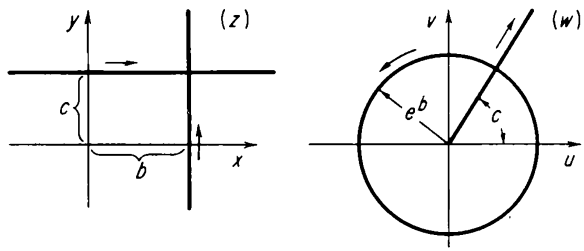

FIGURE 9.3

Next consider the line

$$z = t + ic, \tag{9.28}$$

parallel to the real axis. Then the image of (9.28) under the mapping $w = e^z$ is the curve

$$w = e^t(\cos c + i \sin c), \tag{9.29}$$

i.e., w traces out a ray of slope $\tan c$ emanating from the origin. Moreover, as z describes the line (9.28) once in such a way that t, the abscissa of z, increases continuously from $-\infty$ to $+\infty$, w describes the ray (9.29) once in such a way that the distance of w from the origin increases continuously from 0 to ∞ (of course, the limits 0 and ∞ are excluded, since $|w| = e^t$). Thus, *under the mapping $w = e^z$, a family of lines parallel to the imaginary axis is transformed into a family of concentric circles with the origin as center, and a family of lines parallel to the real axis is transformed into a family of rays emanating from the origin.*

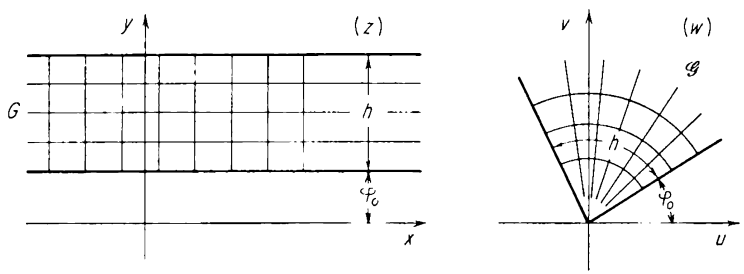

FIGURE 9.4

Now consider the domain G consisting of all points z such that

$$\varphi_0 < \operatorname{Im} z < \varphi_1,$$

where $\varphi_1 - \varphi_0 = h$; such a domain will be called an (*open*) *strip* of width h. Suppose $0 < h < 2\pi$, and let $\mathcal{G}$ be the image of G under the mapping $w = e^z$. It follows from the considerations just given that $\mathcal{G}$ is the interior of the angle of h radians with vertex at the origin, formed by the rays

$$\operatorname{Arg} w = \varphi_0 + 2k\pi, \quad \operatorname{Arg} w = \varphi_1 + 2k\pi \qquad (k = 0, \pm 1, \pm 2, \ldots)$$

(see Figure 9.4). Moreover, the correspondence between the domains G and $\mathcal{G}$ under the mapping $w = e^z$ is *one-to-one*. To see this, we recall that the inverse images of a point $w \in \mathcal{G}$ are all of the form (9.25), and hence differ only in the values of their imaginary parts. In fact, any two points (9.25) lie on a line parallel to the imaginary axis, and the distance between them is an integral multiple of 2π. However, by assumption, the width h of our strip does not exceed 2π, and G can contain only one inverse image of the point w, i.e., not only is $w = f(z)$ a single-valued function on G, but its inverse $z = f^{-1}(w)$ is a single-valued function on $\mathcal{G} = f(G)$. Thus, *the exponential function $w = e^z$ is a one-to-one conformal mapping of an open strip of width $h \leqslant 2\pi$ with sides parallel to the real axis onto the interior of an angle of h radians with vertex at the origin.*

Next consider a straight line with equation

$$z = (1 + i\alpha)t + ib \qquad (-\infty < t < \infty), \tag{9.30}$$

which is not parallel to one of the coordinate axes. Here $\alpha \neq 0$ is the slope of the line (9.30), and b is its y-intercept. The image of (9.30) under the mapping $w = e^z$ is the curve

$$w = \exp[t + i(\alpha t + b)] = e^t[\cos(\alpha t + b) + i \sin(\alpha t + b)].$$

Therefore

$$|w| = r = e^t, \qquad \varphi = \operatorname{Arg} w = \alpha t + b + 2k\pi,$$

and eliminating the parameter t, we obtain

$$r = \exp[(\varphi - b - 2k\pi)/\alpha]. \tag{9.31}$$

If we set $\theta = \varphi - 2k\pi$, (9.31) becomes

$$r = ce^{\theta/\alpha}, \tag{9.32}$$

where $c = e^{-b/\alpha}$. This is the equation (in polar form) of a *logarithmic spiral*. Since the mapping $w = e^z$ is conformal, and since (9.32) is the image of the line (9.30) intersecting all lines parallel to the real axis at the same angle $\arctan \alpha$, it follows that the logarithmic spiral intersects the images of all these lines, i.e., all rays emanating from the origin, at the same angle $\arctan \alpha$, a property which characterizes the logarithmic spiral (see Figure 9.5).

Remark. As is well known,[8]

$$e^x = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n \tag{9.33}$$

for real x. Using (9.33), we can easily show that

$$e^z = \lim_{n \to \infty} \left(1 + \frac{z}{n}\right)^n \tag{9.34}$$

for complex z. In fact, writing

$$z_n = \left(1 + \frac{z}{n}\right)^n,$$

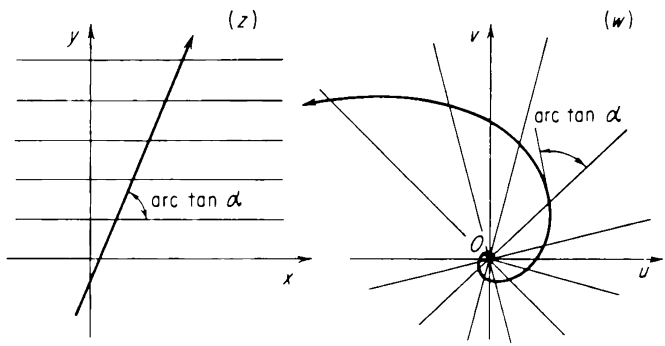

FIGURE 9.5

we have

$$|z_n| = \left|1 + \frac{x}{n} + i\frac{y}{n}\right|^n = \left[\left(1 + \frac{x}{n}\right)^2 + \frac{y^2}{n^2}\right]^{n/2},$$

$$\arg z_n = n \arctan \frac{y/n}{1 + (x/n)}.$$

Therefore

$$\lim_{n \to \infty} |z_n| = \lim_{n \to \infty} \left(1 + \frac{2x}{n}\right)^{n/2} = e^x,$$

where we drop $(x^2 + y^2)/n^2$ in comparison to $2x/n$ and use (9.33). Moreover, replacing small angles by their tangents, we see that

$$\lim_{n \to \infty} \arg z_n = \lim_{n \to \infty} \frac{n(y/n)}{1 + (x/n)} = y.$$

---

[8] See e.g., R. Courant, *Differential and Integral Calculus, Vol. I*, second edition (translated by E. J. McShane), Interscience Publishers, Inc., New York (1959), p. 175.

But then, according to p. 34,

$$\begin{aligned}\lim_{n\to\infty}\left(1+\frac{z}{n}\right)^n &= \lim_{n\to\infty} z_n \\ &= \lim_{n\to\infty}|z_n|\,[\cos(\lim_{n\to\infty}\arg z_n)+i\sin(\lim_{n\to\infty}\arg z_n)] \\ &= e^x(\cos y + i\sin y),\end{aligned}$$

and comparison of this result with (9.22) proves (9.34).

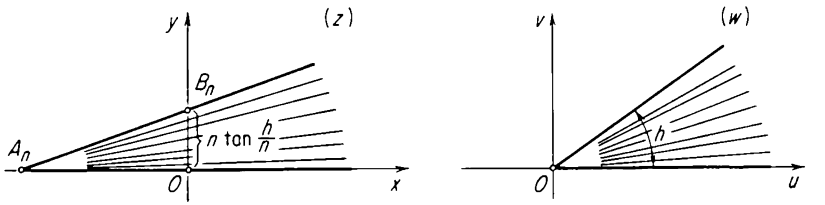

FIGURE 9.6

Formula (9.34) shows the connection between the mapping $w = e^z$ and the mapping $w = (z - a)^n$ studied in Sec. 37. We note that e^z is the limit as $n \to \infty$ of the mapping

$$w = \left(1 + \frac{z}{n}\right)^n = \frac{1}{n^n}[z - (-n)]^n, \tag{9.35}$$

which, as we know, maps the interior of an angle of h/n radians ($0 < h \leqslant 2\pi$) with vertex at the point $A_n = (-n, 0)$ and sides consisting of the rays

$$x \geqslant -n, \qquad y = 0$$

and

$$\operatorname{Arg}(z + n) = \frac{h}{n} + 2k\pi,$$

onto the interior of an angle of h radians with vertex at the origin and sides consisting of the rays

$$\operatorname{Arg} w = 0, \qquad \operatorname{Arg} w = h + 2m\pi.$$

As $n \to \infty$, the vertex A_n approaches infinity along the negative real axis and the length of the segment $\overline{OB_n}$ (see Figure 9.6) approaches

$$\lim_{n\to\infty} n\tan\frac{h}{n} = h,$$

so that the limiting position of the ray $\overline{A_nB_n}$ is the line $y = h$, which together with the real axis forms the boundary of a strip of width h. Moreover, as $n \to \infty$, the rays emanating from the vertex A_n approach lines parallel to the real axis and the arcs of circles with A_n as center approach perpendiculars to

the real axis lying inside the strip. In other words, in the limit as $n \to \infty$, the effect of the mapping (9.35) is exactly the same as that of the mapping $w = e^z$.

40. Some Functions Related to the Exponential

According to the formulas

$$e^{ix} = \cos x + i \sin x, \qquad e^{-ix} = \cos x - i \sin x, \tag{9.36}$$

we have

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}, \qquad \sin x = \frac{e^{ix} - e^{-ix}}{2i} \tag{9.37}$$

for arbitrary real x. If z is an arbitrary finite complex number, we define two (entire) *trigonometric functions* $\cos z$ and $\sin z$, called the *cosine* and *sine*,[9] by simply changing x to z everywhere in (9.37):

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}, \qquad \sin z = \frac{e^{iz} - e^{-iz}}{2i}. \tag{9.38}$$

This seems quite natural, since the functions $\cos z$ and $\sin z$ are obviously analytic for all z, and reduce to the familiar functions $\cos x$ and $\sin x$ when $z = x$ is real. It follows from the definitions (9.38) that $\cos z$ is even and $\sin z$ is odd, i.e., that

$$\cos(-z) = \cos z, \qquad \sin(-z) = -\sin z.$$

Moreover, (9.38) implies the formulas

$$e^{iz} = \cos z + i \sin z, \qquad e^{-iz} = \cos z - i \sin z, \tag{9.39}$$

which generalize (9.36).

The functions $\cos z$ and $\sin z$ are both periodic with period 2π, since changing z to $z + 2\pi$ in (9.38) amounts to multiplying the exponentials by $e^{\pm 2\pi i} = 1$. Actually, 2π is the fundamental period of $\cos z$ and $\sin z$, i.e., any other period is an integral multiple of 2π, as we now verify for $\cos z$. If ω is any period of $\cos z$, then

$$\cos(z + \omega) = \cos z,$$

and hence, setting $z = \pi/2$, we obtain

$$\cos\left(\omega + \frac{\pi}{2}\right) = 0.$$

But this implies

$$\exp\left[i\left(\omega + \frac{\pi}{2}\right)\right] + \exp\left[-i\left(\omega + \frac{\pi}{2}\right)\right] = 0,$$

or

$$\exp[i(2\omega + \pi)] = -1.$$

[9] Trigonometric functions of a more general nature are discussed in Sec. 52.

Therefore, according to formula (9.25),

$$i(2\omega + \pi) = \ln|-1| + i\,\mathrm{Arg}\,(-1) = i(\pi + 2k\pi),$$

so that

$$\omega = 2k\pi,$$

as asserted. Similarly, it can easily be verified that 2π is the fundamental period of $\sin z$.

Next we derive *addition theorems* for the functions $\cos z$ and $\sin z$, i.e., formulas relating the quantities $\cos(z_1 + z_2)$ and $\sin(z_1 + z_2)$ to the quantities $\cos z_1$, $\sin z_1$, $\cos z_2$ and $\sin z_2$, where z_1 and z_2 are arbitrary complex numbers. As might be expected, the required relations are immediate consequences of the addition theorem

$$\exp(z_1 + z_2) = \exp z_1 \exp z_2$$

for the exponential. In fact, replacing z by $z_1 + z_2$ in the formulas (9.39), we find that

$$\begin{aligned}\cos(z_1 + z_2) + i\sin(z_1 + z_2) &= \exp[i(z_1 + z_2)] = \exp(iz_1)\exp(iz_2)\\ &= (\cos z_1 + i\sin z_1)(\cos z_2 + i\sin z_2)\\ &= (\cos z_1 \cos z_2 - \sin z_1 \sin z_2) + i(\sin z_1 \cos z_2 + \cos z_1 \sin z_2)\end{aligned} \tag{9.40}$$

and

$$\begin{aligned}\cos(z_1 + z_2) - i\sin(z_1 + z_2) &= \exp[-i(z_1 + z_2)] = \exp(-iz_1)\exp(-iz_2)\\ &= (\cos z_1 - i\sin z_1)(\cos z_2 - i\sin z_2)\\ &= (\cos z_1 \cos z_2 - \sin z_1 \sin z_2) - i(\sin z_1 \cos z_2 + \cos z_1 \sin z_2).\end{aligned} \tag{9.41}$$

First adding (9.41) to (9.40), and then subtracting (9.41) from (9.40), we obtain the addition theorems

$$\begin{aligned}\cos(z_1 + z_2) &= \cos z_1 \cos z_2 - \sin z_1 \sin z_2,\\ \sin(z_1 + z_2) &= \sin z_1 \cos z_2 + \cos z_1 \sin z_2,\end{aligned} \tag{9.42}$$

which are basic in the theory of trigonometric functions. In particular, the so-called *reduction formulas* are implicit in (9.42). For example, setting $z_1 = z$, $z_2 = \pi/2$ in (9.42) gives

$$\cos\left(z + \frac{\pi}{2}\right) = \cos z \cos\frac{\pi}{2} - \sin z \sin\frac{\pi}{2} = -\sin z,$$

$$\sin\left(z + \frac{\pi}{2}\right) = \sin z \cos\frac{\pi}{2} + \cos z \sin\frac{\pi}{2} = \cos z,$$

setting $z_1 = z$, $z_2 = \pi$, gives

$$\begin{aligned}\cos(z + \pi) &= -\cos z,\\ \sin(z + \pi) &= -\sin z,\end{aligned}$$

and so on. Moreover, substituting $z_1 = z$, $z_2 = -z$ into the first of the formulas (9.42), we obtain the following basic relation between $\cos z$ and $\sin z$:

$$\cos^2 z + \sin^2 z = 1. \tag{9.43}$$

Thus we see that all the familiar trigonometric formulas involving cosines and sines of real arguments remain valid in the complex domain. However, we cannot infer from (9.43) that $|\cos z| \leqslant 1$ and $|\sin z| \leqslant 1$, since $\cos^2 z$ and $\sin^2 z$ are generally not nonnegative real numbers.

Closely related to the trigonometric functions $\cos z$ and $\sin z$ are the *hyperbolic functions* $\cosh z$ and $\sinh z$, called the *hyperbolic cosine* and the *hyperbolic sine*, respectively, and defined by the formulas

$$\cosh z = \frac{e^z + e^{-z}}{2}, \qquad \sinh z = \frac{e^z - e^{-z}}{2}. \tag{9.44}$$

Clearly, the functions (9.44) are analytic for all finite z, and reduce to the familiar hyperbolic functions $\cosh x$ and $\sinh x$ when $z = x$ is real. It will be recalled that the function $\cosh x$ is even and decreases monotonically from $+\infty$ to 1 as x increases from $-\infty$ to 0, and then increases monotonically from 1 to $+\infty$ again as x increases from 0 to $+\infty$. On the other hand, the function $\sinh x$ is odd, and increases monotonically from $-\infty$ to $+\infty$ as x increases from $-\infty$ to $+\infty$. Comparing (9.44) and (9.37), we see that the trigonometric and hyperbolic functions are connected by the relations

$$\cosh z = \cos (iz), \qquad \sinh z = -i \sin (iz). \tag{9.45}$$

In particular, it follows that

$$\cosh^2 z - \sinh^2 z = \cos^2 (iz) + \sin^2 (iz) = 1. \tag{9.46}$$

Next we derive expressions for the real and imaginary parts, and also the moduli, of the functions $\cos z$ and $\sin z$. Setting $z = x + iy$, and using formulas (9.42) and (9.45), we find that

$$\begin{aligned} \cos (x + iy) &= \cos x \cos (iy) - \sin x \sin (iy) = \cos x \cosh y - i \sin x \sinh y, \\ \sin (x + iy) &= \sin x \cos (iy) + \cos x \sin (iy) = \sin x \cosh y + i \cos x \sinh y, \end{aligned} \tag{9.47}$$

which imply

$$\begin{aligned} &\operatorname{Re} \cos (x + iy) = \cos x \cosh y, \qquad &&\operatorname{Im} \cos (x + iy) = -\sin x \sinh y, \\ &\operatorname{Re} \sin (x + iy) = \sin x \cosh y, \qquad &&\operatorname{Im} \sin (x + iy) = \cos x \sinh y. \end{aligned}$$

Thus for the moduli of $\cos z$ and $\sin z$ we obtain the expressions

$$\begin{aligned} |\cos z| &= \sqrt{(\cos x \cosh y)^2 + (\sin x \sinh y)^2} \\ &= \sqrt{\cosh^2 y (1 - \sin^2 x) + \sin^2 x \sinh^2 y} = \sqrt{\cosh^2 y - \sin^2 x}, \\ |\sin z| &= \sqrt{(\sin x \cosh y)^2 + (\cos x \sinh y)^2} \\ &= \sqrt{\sin^2 x (1 + \sinh^2 y) + \cos^2 x \sinh^2 y} = \sqrt{\sinh^2 y + \sin^2 x}, \end{aligned}$$

which in turn imply the inequalities

$$\begin{aligned} |\sinh y| &= \sqrt{\cosh^2 y - 1} \leqslant |\cos z| \leqslant \cosh y, \\ |\sinh y| &\leqslant |\sin z| \leqslant \sqrt{\sinh^2 y + 1} = \cosh y. \end{aligned} \tag{9.48}$$

Thus we see that the moduli of the functions cos z and sin z increase without limit as $|y|$ increases, i.e., as the distance from z to the real axis increases, but this increase is no faster than that of cosh y and no slower than that of sinh y. In Figure 9.7 we plot the surface $\mu = |\sin z|$, called the *modular surface* of the function sin z.[10]

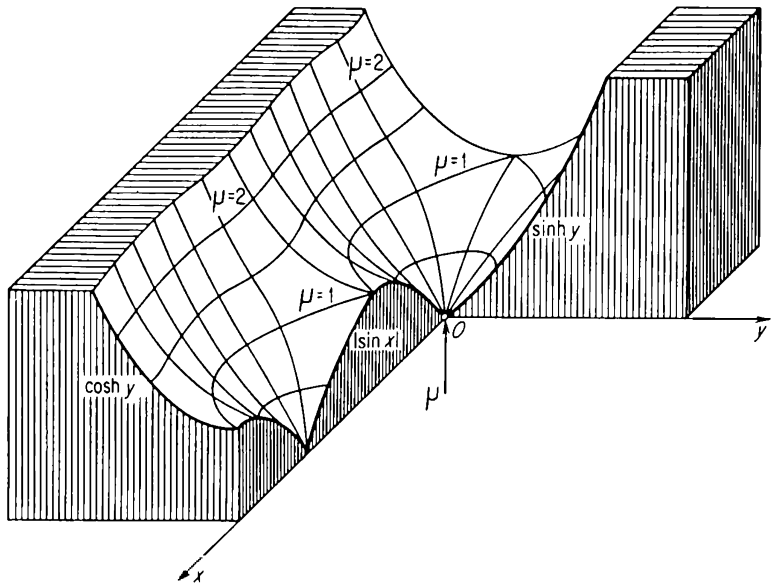

FIGURE 9.7

Since sinh $y \neq 0$ for $y \neq 0$, it follows from the inequalities (9.48) that cos z and sin z cannot vanish except on the real axis, i.e., the equations cos $z = 0$ and sin $z = 0$ have no imaginary roots. Therefore the roots of the equation cos $z = 0$ reduce to

$$z = (2k - 1)\frac{\pi}{2} \qquad (k = 0, \pm 1, \pm 2, \ldots),$$

while those of the equation sin $z = 0$ reduce to

$$z = k\pi \qquad (k = 0, \pm 1, \pm 2, \ldots),$$

which are just the roots familiar from elementary trigonometry. Finally, differentiating the trigonometric and hyperbolic functions, we obtain

$$(\cos z)' = \frac{1}{2}(e^{iz} + e^{-iz})' = \frac{i}{2}(e^{iz} - e^{-iz}) = -\sin z,$$

and similarly,

$$(\sin z)' = \cos z, \quad (\cosh z)' = \sinh z, \quad (\sinh z)' = \cosh z.$$

[10] The unlabelled curves correspond to the equation arg sin z = const.

41. The Mapping $w = \cos z$

In studying the geometric behavior of the trigonometric functions, we shall confine ourselves to the mapping $w = \cos z$, since the mapping $w = \sin z$ can be written in the form

$$w = -\cos\left(z + \frac{\pi}{2}\right),$$

and hence reduces to a shift

$$z_1 = z + \frac{\pi}{2}$$

of the z-plane in the direction of the positive real axis, followed first by the mapping $z_2 = \cos z_1$ and then by the mapping $w = -z_2$, corresponding to a rotation of the whole z_2-plane through the angle π about the origin.

First we consider the inverse images of the point w under the mapping $w = \cos z$, i.e., the roots of the equation

$$\cos z = w, \tag{9.49}$$

where w is an arbitrary finite complex number. Substituting (9.38) for $\cos z$ and writing

$$e^{iz} = t \tag{9.50}$$

for brevity, we find that t satisfies the equation

$$\frac{1}{2}\left(t + \frac{1}{t}\right) = w$$

or

$$t^2 - 2wt + 1 = 0, \tag{9.51}$$

with solutions

$$t_j = w + \sqrt{w^2 - 1} \qquad (j = 1, 2). \tag{9.52}$$

In (9.52) we do not write $\pm$ in front of the radical, since the square root is already understood to have two values (see Sec. 6). Obviously, the product of the two numbers t_1 and t_2 equals 1, and hence $t_1 \neq 0$, $t_2 \neq 0$. If we denote one of these numbers by τ and the other by $1/\tau$, (9.50) leads to two equations for determining z:

$$e^{iz} = \tau\ (\neq 0), \qquad e^{iz} = \frac{1}{\tau}\ (\neq 0).$$

According to Sec. 39, each of these equations has infinitely many solutions, found by using formula (9.25), i.e.,

$$iz' = \ln|\tau| + i \operatorname{Arg} \tau, \qquad iz'' = \ln\left|\frac{1}{\tau}\right| + i \operatorname{Arg} \frac{1}{\tau}$$

or

$$z' = \operatorname{Arg} \tau - i \ln|\tau|, \qquad z'' = -(\operatorname{Arg} \tau - i \ln|\tau|).$$

Therefore the solutions of (9.49) consist of two infinite sets of points z'_n, z''_n $(n = 0, \pm 1, \pm 2, \ldots)$ lying on the lines $y = \pm \ln|\tau|$ parallel to the real axis, where each adjacent pair of points z'_n is separated by the distance 2π, and the same is true of the points z''_n. For each point z'_n on the line $y = -\ln|\tau|$ there is a point z''_n on the line $y = \ln|\tau|$ which is symmetric to z'_n with respect to the origin (see Figure 9.8). For $w = \pm 1$ the roots τ and $1/\tau$ of equation (9.51) become ± 1, and then both lines coincide with the real axis, and both sets of points z'_n, z''_n also coincide. Thus, in any case, equation (9.49) has infinitely many solutions. It follows that 1) the function $w = \cos z$ maps the finite z-plane *onto* the finite w-plane, and 2) each point w has infinitely many inverse images in the z-plane. Moreover, the mapping is conformal at all points where

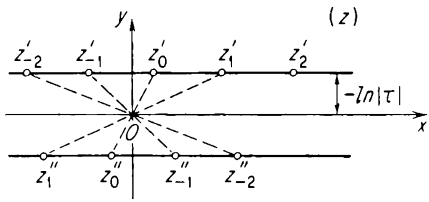

FIGURE 9.8

$$(\cos z)' = -\sin z \neq 0,$$

i.e., for

$$z \neq k\pi \qquad (k = 0, \pm 1, \pm 2, \ldots).$$

Now consider the effect of the mapping $w = \cos z$ on straight lines parallel to one of the coordinate axes. Every line l parallel to the imaginary axis has an equation of the form

$$z = b + it \qquad (-\infty < t < \infty),$$

where b is a real number. The image of l under the mapping $w = \cos z$ is the curve L with equation

$$w = u + iv = \cos(b + it) = \cos b \cosh t - i \sin b \sinh t \quad (9.53)$$

[cf. (9.47)]. If $b = k\pi$, where k is an integer, (9.53) reduces to

$$w = \cos k\pi \cosh t = (-1)^k \cosh t,$$

i.e., w describes the ray $u \geqslant 1$, $v = 0$ twice if k is even, or the ray $u \leqslant -1$, $v = 0$ twice if k is odd (these rays are both parts of the real axis). On the other hand, if $b = (2k - 1)\pi/2$, (9.53) reduces to

$$w = (-1)^k i \sinh t,$$

i.e., w describes the whole imaginary axis once in the direction of increasing u if k is even, or once in the direction of decreasing u if k is odd. Finally, if b is not an integral multiple of $\pi/2$, we separate real and imaginary parts of (9.53), obtaining the following parametric equations for L:

$$u = \cos b \cosh t, \qquad v = -\sin b \sinh t. \tag{9.54}$$

Elimination of the parameter t leads to

$$\frac{u^2}{\cos^2 b} - \frac{v^2}{\sin^2 b} = 1, \tag{9.55}$$

where $\cos b \neq 0$, $\sin b \neq 0$. This is the equation of a hyperbola with semiaxes $|\cos b|$ and $|\sin b|$, and with foci at the points $w = \pm 1$.

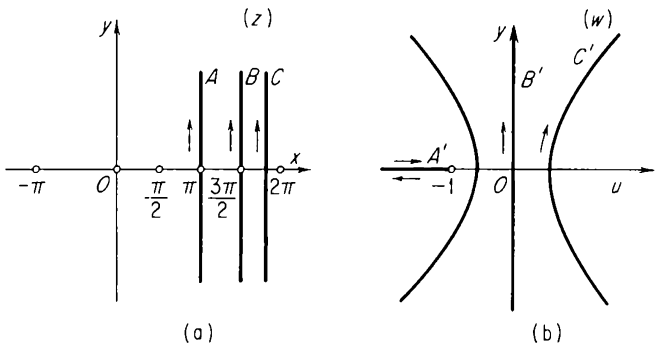

FIGURE 9.9

Remark. It should not be thought that the curve L with parametric representation (9.54) consists of both branches of the hyperbola with equation (9.55). In fact, it follows from (9.54) that u always has the same sign as $\cos b$, and only v changes sign, varying monotonically from $-\infty$ to $+\infty$ as t increases if $\sin b < 0$, or from $+\infty$ to $-\infty$ if $\sin b > 0$. In other words, L actually consists of just one of the branches of the hyperbola (9.55), i.e., the right-hand branch if $\cos b > 0$, or the left-hand branch if $\cos b < 0$. In every case, the mapping of the line l onto the curve L is one-to-one, and in fact, there is a one-to-one correspondence between the points of each of the half-lines into which l is divided by the real axis and the points of one of the half-branches into which L is divided by its vertex.

Example. Consider the three lines

$$A) \quad x = \pi,$$

$$B) \quad x = \frac{3\pi}{2},$$

$$C) \quad x = b \quad \left(\frac{3\pi}{2} < b < 2\pi\right),$$

shown in Figure 9.9(a). The images of these three lines are shown in Figure 9.9(b) and are A') a ray along the negative real axis, B') the imaginary axis itself, and C') the right-hand branch of a hyperbola.

Next suppose z describes a line λ parallel to the real axis, with equation

$$z = t + ic \qquad (-\infty < t < \infty),$$

where c is a real number. The image of λ under the mapping $w = \cos z$ is the curve Λ with equation

$$w = u + iv = \cos(t + ic) = \cos t \cosh c - i \sin t \sinh c. \tag{9.56}$$

If $c = 0$, λ is the real axis and Λ is the curve with equation

$$w = \cos t \qquad (-\infty < t < \infty),$$

i.e., w describes the segment $-1 \leqslant u \leqslant 1$ of the real axis infinitely many times, in fact twice every time z goes a distance 2π along λ. If $c \neq 0$, we

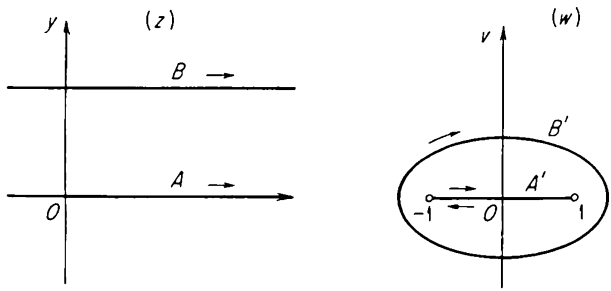

FIGURE 9.10

separate real and imaginary parts of (9.56), obtaining the following parametric equations for Λ:

$$u = \cosh c \cos t, \qquad v = -\sinh c \sin t. \tag{9.57}$$

Elimination of the parameter t gives

$$\frac{u^2}{\cosh^2 c} + \frac{v^2}{\sinh^2 c} = 1,$$

where $\cosh c \neq 0$, $\sinh c \neq 0$. This is the equation of an ellipse with semiaxes $|\cosh c|$ and $|\sinh c|$, and with foci at the points $w = \pm 1$. It follows from the representation (9.57) of the curve Λ that the point w describes the ellipse infinitely many times in the same direction, where each circuit around the ellipse corresponds to a displacement of the point z a distance 2π along the line λ. The two lines

$$A) \quad y = 0,$$
$$B) \quad y = c \neq 0,$$

and their images A' and B' in the w-plane are shown in Figure 9.10.

To summarize, *the mapping* $w = \cos z$ *transforms the two one-parameter families of straight lines parallel to the coordinate axes, but distinct from the lines* $x = k\pi$ $(k = 0, \pm 1, \pm 2, \ldots)$ *or* $y = 0$, *into a one-parameter family of confocal hyperbolas and a one-parameter family of confocal ellipses, with common foci at the points* $w = \pm 1$. Moreover, since the mapping is conformal at all points of the z-plane except at the points

$$z = k\pi \qquad (k = 0, \pm 1, \pm 2, \ldots), \tag{9.58}$$

whose images are precisely the foci $w = \pm 1$, and since the two families of lines in the z-plane form an orthogonal system (see p. 134), it follows that the family of hyperbolas and the family of ellipses in the w-plane also form an orthogonal system.

42. The Image of a Half-Strip under $w = \cos z$

We now pose the problem of finding a domain G in the z-plane on which the function $w = \cos z$ is one-to-one. The choice of such a domain can be made in many ways. It is only necessary to make sure that G does not contain two or more inverse images of the same point w. For example,

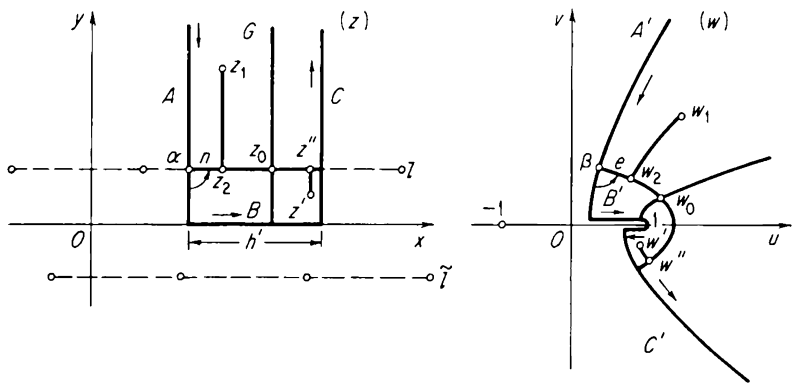

FIGURE 9.11

suppose G is the open half-strip of width h $(0 < h \leqslant 2\pi)$ whose sides are parallel to the imaginary axis and whose base is a segment of the real axis [see Figure 9.11(a)]. Then $w = \cos z$ is a one-to-one function on G. In fact, as we know from the preceding section, if $z_0 \in G$ and $w_0 = \cos z_0$, all the other inverse images of w_0 must lie either on the line l parallel to the real axis and passing through the point z_0, or on the line $\tilde{l}$ which is symmetric to l with respect to the real axis. But the distance from z_0 to any other inverse image of w_0 lying on l is an integral multiple of 2π, and since the width of the half-strip does not exceed 2π, none of the other inverse images on l can belong to either G or its boundary. Moreover, G and the line $\tilde{l}$ obviously

have no points in common. It follows that the function $w = \cos z$ is a one-to-one conformal mapping of G onto some set of points $\mathcal{G}$ in the w-plane [note that none of the points (9.58) belongs to G].

To construct $\mathcal{G}$, we proceed as follows: Let the point $z = x + iy$ trace out the boundary γ of the half-strip in the way indicated in Figure 9.11(a), i.e., let z consecutively traverse the following three curves in the z-plane:

A) The left-hand side of the half-strip, in the direction of decreasing y;
B) The base of the half-strip, in the direction of increasing x;
C) The right-hand side of the half-strip, in the direction of increasing y.

This causes the image point $w = u + iv = \cos z$ to consecutively traverse the following three curves in the w-plane, as indicated in Figure 9.11(b):[11]

A') Half of a branch of a hyperbola, in the direction of decreasing v;
B') Part of the segment $-1 \leqslant u \leqslant 1$ of the real axis, first in the direction of increasing u, and then in the direction of decreasing u;
C') Half of the branch of another hyperbola, in the direction of decreasing v.

The union of the three curves A', B' and C', which we denote by Γ, is the image (under $w = \cos z$) of γ, the boundary of G. Clearly, Γ is a closed Jordan curve, passing through the point at infinity. Therefore, Γ divides the extended w-plane into two disjoint domains D_1 and D_2, with Γ as their common boundary (cf. Secs. 18, 22). According to Theorem 6.3, one of these two domains is the set $\mathcal{G}$, the image of G. To ascertain whether $D_1 = \mathcal{G}$ or $D_2 = \mathcal{G}$, we choose a point $z_0 \in G$ and find its image $w_0 = \cos z_0$. Clearly $w_0 \notin \Gamma$, since otherwise one of the inverse images of w_0 would belong to G and another to γ, which, as we have just seen, is impossible. Therefore, w_0 belongs to either D_1 or D_2, and the domain containing w_0 is $\mathcal{G}$ (see the remark on p. 100).

Remark 1. In the case of an elementary mapping like $w = \cos z$, we can use the detailed geometry of the mapping to prove that $\mathcal{G}$ coincides with one of the domains D_1 and D_2 into which Γ divides the w-plane, thereby avoiding the topological considerations of Chap. 6. Let z_0 be an arbitrary (but fixed) point of G, and let $w_0 = \cos z_0$ be its image. As already noted, one of the domains D_1 and D_2 contains w_0; let this domain be denoted simply by D. First we show that the image w_1 of any point $z_1 \in G$ different from z_0 belongs to D, so that $\mathcal{G} \subset D$. Through the point z_0 draw the line l parallel to the real axis, and through the point z_1 draw the line parallel to the imaginary axis.

---

[11] Of course, the specific appearance of the curves A', B' and C', and the directions in which they are traversed, depends on the location and width of the half-strip in the z-plane. In this respect, Figure 9.11 represents only one possible case (see Remark 3, p. 157 and Problem 9.16). For example, if the half-strip is of width 2π, the curve B' will consist of the whole segment $-1 \leqslant u \leqslant 1$ traced out *twice*.

Let z_2 be the point of intersection of these two lines [see Figure 9.11(a)]. Then the polygonal curve[12]

$$\lambda = \overline{z_0 z_2} + \overline{z_2 z_1}$$

joins z_0 to z_1, and moreover $\lambda \subset G$. (Either $z_0 = z_2$ or $z_2 = z_1$ is possible, but not both.) The image of λ under the mapping $w = \cos z$ is a curve

$$\Lambda = \widehat{w_0 w_2} + \widehat{w_2 w_1},$$

where $\widehat{w_0 w_2}$ is an arc of an ellipse and $\widehat{w_2 w_1}$ is an arc of a hyperbola, both with foci ± 1 ($w_1 = \cos z_1$, $w_2 = \cos z_2$). Clearly $\Lambda \cap \Gamma = 0$, since $\lambda \cap \gamma = 0$ and no pair of points, one in G and the other on its boundary, can have the same image. Therefore $\Lambda \subset D$, and hence $w_1 \in D$, since Λ connects w_0 to w_1, i.e., $\mathscr{G} \subset D$, as asserted.

Next, let w' be a point of D different from w_0. We now show that w' has a (unique) inverse image belonging to G, so that $\mathscr{G} \supset D$. Since $w' \notin \Gamma$, there is a unique ellipse with foci ± 1 passing through w_0, and a unique hyperbola with foci ± 1 passing through w'. Let w'' be the point of intersection of these two curves [see Figure 9.11(b)]. Then the curve

$$\Lambda = \widehat{w_0 w''} + \widehat{w'' w'}$$

joins w_0 to w', and moreover $\Lambda \subset D$. (Either $w_0 = w''$ or $w'' = w'$ is possible, but not both.) Any inverse image of Λ under $w = \cos z$ is a polygonal curve

$$\lambda = \overline{z_0 z''} + \overline{z'' z'},$$

where $\overline{z_0 z''}$ is a segment of a line parallel to the real axis, $\overline{z'' z'}$ is a segment of a line parallel to the imaginary axis, and z', z'' are any two points such that $w' = \cos z'$, $w'' = \cos z''$. Obviously $\lambda \cap \gamma = 0$, since $\Lambda \cap \Gamma = 0$. Therefore $\lambda \subset G$, and hence $z' \in G$, since λ connects z_0 to z', i.e., $\mathscr{G} \subset D$, as asserted, and the proof is complete.[13]

Remark 2. We now indicate another way of deciding which of the two domains D_1 and D_2 with boundary Γ coincides with $\mathscr{G}$. This method does not require knowing the image w_0 of a point $z_0 \in G$, and goes as follows: As

---

[12] By $\widehat{ab}$ we mean a curve joining a to b (the curve in question is always apparent from the context). In the case where the curve is the line segment joining a to b, we write $\overline{ab}$ instead of ab. If γ and δ are two curves such that the final point of γ coincides with the initial point of δ, then by $\gamma + \delta$ we mean the *oriented* curve obtained by first going along γ from its initial point to its final point and then along δ from its initial point to its final point. We use the symbol $+$ instead of $\cup$ to emphasize that $\gamma + \delta$ is traversed in the direction indicated. It should be noted that $\widehat{z_0 z_2} \cup \widehat{z_2 z_1} = \widehat{z_2 z_1} \cup \widehat{z_0 z_2}$, but $\widehat{z_2 z_1} + \widehat{z_0 z_2}$ is meaningless.

[13] We can now state that z', z'' and hence $\overline{z_0 z''}$, $\overline{z'' z'}$ are unique, since $z' \in G$, $z'' \in G$.

a moving point z traces out the boundary γ of the domain G in a certain direction, the image point w traces out the boundary Γ of the domain $\mathscr{G}$ in a corresponding direction. For example, the direction along γ indicated by the arrows in Figure 9.11(a) corresponds to the direction along Γ indicated by the arrows in Figure 9.11(b), and if one of these directions were reversed, the other would also have to be reversed. Suppose an observer (labeled *I*) follows the moving point z along γ, and suppose another observer (labeled *II*) follows the moving image point $w = \cos z$ along Γ. Then if *I* finds G on his left, *II* finds $\mathscr{G}$ on his left, while if *I* finds G on his right, *II* finds $\mathscr{G}$ on his right. To prove this, suppose *I* moves along the side A of the half-strip G in the direction shown in Figure 9.11(a) and hence finds G on his left. From some point $\alpha \in A$ ($\operatorname{Im} \alpha > 0$) we draw a segment n of the normal to A pointing into G, i.e., a certain line segment parallel to the real axis. Obviously $\mathscr{G}$ must contain the image of this segment, which is a certain elliptical arc e drawn from the point $\beta = \cos \alpha$ belonging to A', as shown in Figure 9.11(b). In other words, $\mathscr{G}$ is uniquely characterized as the domain with boundary Γ into which e points. The assertion that *I* finds G on his left is equivalent to the assertion that in order to enter G along the segment n at the point α, *I* must make a "left turn," i.e., a counterclockwise rotation through 90°. But $w = \cos z$ has a nonzero derivative at α, and hence is a conformal mapping of the first kind which preserves not only angles but also the directions in which they are measured, as discussed in Sec. 31. Therefore, in order to enter $\mathscr{G}$ along the arc e at the point β, *II* must also make a left turn, i.e., *II* finds $\mathscr{G}$ on his left, as required.

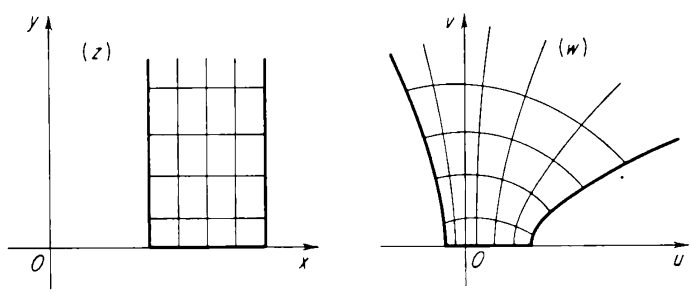

FIGURE 9.12

Remark 3. Once again (cf. footnote 11), we mention that the specific appearance of the domain $\mathscr{G}$ depends on the location and width of the half-strip G. The case where the base of G is contained in an interval of the form $[k\pi, (k + 1)\pi]$ is shown in Figure 9.12. The situation illustrated by Figure 9.11 corresponds to the case where the base of G is an interval containing a point of the form $k\pi$ in its interior. In this case, the boundary Γ is "folded over" at either $w = -1$ or $w = +1$ (or both).

PROBLEMS

9.1. Given two polynomials $f(z)$ and $g(z)$, prove that there exist uniquely defined polynomials $q(z)$ and $r(z)$ such that

$$f(z) = q(z)g(z) + r(z),$$

where the degree of $r(z)$ is less than the degree of $g(z)$.

Hint. Suppose $f(z) = a_0 + a_1 z + \cdots + a_m z^m$, $g(z) = b_0 + b_1 z + \cdots + b_n z^n$, where $a_m \neq 0$, $b_n \neq 0$. If $m < n$, the result is trivial. If $m \geqslant n$, form the new polynomial

$$f_1(z) = f(z) - \frac{a_m}{b_n} z^{m-n} g(z).$$

Repeat this argument if necessary.

9.2. Use the result of the preceding problem to prove that a necessary and sufficient condition for a polynomial $f(z)$ to be divisible by $z - a$ without a remainder is that $f(a) = 0$, i.e., that $z = a$ be a zero of $f(z)$.

9.3. Prove that if a polynomial of degree no higher than n has more than n distinct zeros, then the polynomial is identically zero. Prove that if the values of two polynomials of degree no higher than n are the same for more than n different values of z, then the two polynomials are identically equal.

9.4. Prove that the factorization (9.2) is unique.

Hint. Suppose that besides (9.2),

$$P(z) = a_n'(z - \alpha_1')^{k_1'} \cdots (z - \alpha_q')^{k_q'}.$$

Equate the two expressions for $P(z)$, and observe that at least one of the numbers $\alpha_1', \ldots, \alpha_q'$ must equal α_1. Repeat this argument if necessary.

9.5. Prove that a necessary and sufficient condition for $z = a$ to be a zero of order k of the polynomial $f(z)$ is that

$$f(a) = f'(a) = \cdots = f^{(k-1)}(a) = 0, \qquad f^{(k)}(a) \neq 0.$$

Hint. Consider the expansion (9.2).

9.6. Prove that if $z = a + ib$ is a zero of order k of the polynomial $f(z)$ with real coefficients, then $\bar{z} = a - ib$ is also a zero of order k of $f(z)$.

9.7. What are the inverse images under the mapping $w = z^2$ of the lines Re w = const, Im w = const?

9.8. In connection with Sec. 37, verify *directly* that any parabola of the family (9.10) is orthogonal to any parameter of the family (9.11).

9.9. Express the complex numbers ± 1, $\pm i$, $1 \pm i$, $-1 \pm i$ in polar form (see p. 140).

9.10. Find the moduli and principal values of the arguments of the following complex numbers:

a) e^{2+i}; b) e^{2-3i}; c) e^{3+4i}; d) e^{-3-4i}; e) $-ae^{i\varphi}$ $(a > 0, |\varphi| \leqslant \pi)$.

9.11. What is the image under the mapping $w = e^z$ of the strip in the z-plane whose sides consist of the parallel inclined lines

$$y = k(x - a_1), \quad y = k(x - a_2) \qquad (k \neq 0, \quad k \neq \infty)?$$

What is the condition for the mapping to be one-to-one?

9.12. What is the image under the mapping $w = z + e^z$ of the closed strip $-\pi \leqslant y \leqslant \pi$ in the z-plane?

9.13. The function $f(z) = e^{-1/z}$ is defined everywhere except at the point $z = 0$. Prove that

a) $f(z)$ is bounded but not continuous on the region $0 < |z| \leqslant 1$, $|\arg z| \leqslant \frac{1}{2}\pi$;

b) $f(z)$ is continuous but not uniformly continuous on the domain $0 < |z| < 1$, $|\arg z| < \frac{1}{2}\pi$;

c) $f(z)$ is uniformly continuous on the sector $0 < |z| \leqslant 1$, $|\arg z| \leqslant \alpha < \frac{1}{2}\pi$.

Comment. The word *region* is often used as in a), without the precise meaning of the term *domain* (an open connected set).

9.14. Find the real and imaginary parts of the following complex numbers:

a) $\cos(2 + i)$; b) $\sin 2i$; c) $\cosh(2 - i)$; d) $\sinh e^i$.

9.15. What curves are the inverse images under the mapping $w = \cos z$ of the imaginary axis $\operatorname{Im} w = 0$?

9.16. What assumptions about the values of the abscissas of the sides of the half-strip G in the z-plane are implicit in Figure 9.11?

9.17. Give an analysis paralleling that of Secs. 41 and 42 of the mappings

a) $w = \sin z$; b) $w = \cosh z$; c) $w = \sinh z$.

10

ELEMENTARY MEROMORPHIC FUNCTIONS

43. Rational Functions

In the preceding chapter, we became acquainted with some of the simplest members of the class of entire functions. A more general class of functions is the class of meromorphic functions. By a *meromorphic function* we mean a function which can be represented as a ratio of two entire functions.[1] Obviously, every entire function $f(z)$ is meromorphic, since it can be written in the form $f(z)/1$, but the converse is not true, as shown by the function $1/z$ which is meromorphic but not entire, since it becomes infinite for $z = 0$.

The simplest members of the class of meromorphic functions in the full sense of the term (i.e., meromorphic functions which do not reduce to entire functions) are the rational functions. By a *rational function* we mean a function which can be written as a ratio

$$f(z) = \frac{P(z)}{Q(z)} = \frac{a_0 + a_1 z + \cdots + a_m z^m}{b_0 + b_1 z + \cdots + b_n z^n} \qquad (a_m \neq 0, \quad b_n \neq 0) \quad (10.1)$$

of two polynomials $P(z)$ and $Q(z)$, where it is assumed that the fraction $P(z)/Q(z)$ is in *lowest terms*, i.e., that the equations $P(z) = 0$ and $Q(z) = 0$ have no common roots. Let $\alpha_1, \ldots, \alpha_p$ denote the distinct zeros of the polynomial $P(z)$, and let α_s be of order k_s $(s = 1, \ldots, p)$. Similarly, let

[1] The word *meromorphic* stems from the Greek μερος = *fraction* and μορφή = *form*, and means "like a fraction."

$\beta_1, \ldots, \beta_q$ denote the distinct zeros of the polynomial $Q(z)$, and let β_t be of order l_t $(t = 1, \ldots, q)$. Then (10.1) can be written in the form

$$f(z) = \frac{P(z)}{Q(z)} = \frac{a_m(z - \alpha_1)^{k_1} \cdots (z - \alpha_p)^{k_p}}{b_n(z - \beta_1)^{l_1} \cdots (z - \beta_q)^{l_q}}.$$

Obviously

$$k_1 + \cdots + k_p = m, \qquad l_1 + \cdots + l_q = n,$$

and none of the numbers $\alpha_1, \ldots, \alpha_p$ equals any of the numbers $\beta_1, \ldots, \beta_q$, since $f(z)$ is assumed to be in lowest terms. The function $f(z)$ vanishes at each point α_s $(s = 1, \ldots, p)$, called a *zero* (or *zero-point*) *of order* k_s [of $f(z)$], and becomes infinite at each point β_t $(t = 1, \ldots, q)$, called a *pole of order* l_t. The zero α_s is said to be *simple* if $k_s = 1$ and *multiple* if $k_s > 1$; similarly, the pole β_t is said to be *simple* if $l_t = 1$ and *multiple* if $l_t > 1$. It follows from our definition that the zeros of $f(z)$ are poles of $1/f(z)$, while the poles of $f(z)$ are zeros of $1/f(z)$, and moreover that orders are preserved, i.e., a zero of $f(z)$ of a given order becomes a pole of $1/f(z)$ of the same order, and *vice versa.*

To define $f(z)$ when z is the point at infinity, we use the relation

$$f(\infty) = \lim_{z \to \infty} f(z),$$

obtaining

$$\begin{aligned} &1) \quad f(\infty) = 0 \qquad \text{if} \quad m < n; \\ &2) \quad f(\infty) = \frac{a_m}{b_n} \quad \text{if} \quad m = n; \\ &3) \quad f(\infty) = \infty \qquad \text{if} \quad m > n. \end{aligned} \tag{10.2}$$

In the first case, we say that $f(z)$ has a *zero at* ∞, and in the third case we say that $f(z)$ has a *pole at* ∞. To assign a definite order to a zero or pole at ∞, we make the preliminary transformation $\zeta = 1/z$ (cf. Sec. 24), which carries the point $z = \infty$ into the point $\zeta = 0$. Then

$$f^*(\zeta) = f(1/\zeta) = \frac{a_0 + a_1\dfrac{1}{\zeta} + \cdots + a_m\dfrac{1}{\zeta^m}}{b_0 + b_1\dfrac{1}{\zeta} + \cdots + b_n\dfrac{1}{\zeta^n}} = \frac{\zeta^n}{\zeta^m}\frac{a_m + a_{m-1}\zeta + \cdots + a_0\zeta^m}{b_n + b_{n-1}\zeta + \cdots + b_0\zeta^n},$$

and analyzing the three cases (10.2), we obtain the following results:

1. If $m < n$, then

$$f^*(\zeta) = \frac{\zeta^{n-m}(a_m + a_{m-1}\zeta + \cdots + a_0\zeta^m)}{b_n + b_{n-1}\zeta + \cdots + b_0\zeta^n}.$$

Since this rational function has a zero of order $n - m$ at $\zeta = 0$, we say that $f(z)$ has a zero of order $n - m$ at $z = \infty$.

2. If $m = n$, then

$$f^*(\zeta) = \frac{a_m + a_{m-1}\zeta + \cdots + a_0\zeta^m}{b_m + b_{m-1}\zeta + \cdots + b_0\zeta^m}.$$

Since this rational function has neither a zero nor a pole at $\zeta = 0$ [in fact, $f^*(0) = f(\infty) = a_m/b_m$], we say that $f(z)$ has neither a pole nor a zero at $z = \infty$.

3. If $m > n$, then

$$f^*(\zeta) = \frac{a_m + a_{m-1}\zeta + \cdots + a_0\zeta^m}{\zeta^{m-n}(b_n + b_{n-1}\zeta + \cdots + b_0\zeta^n)}.$$

Since this rational function has a pole of order $m - n$ at $\zeta = 0$, we say that $f(z)$ has a pole of order $m - n$ at $z = \infty$.

Remark. It is easy to see that $f(z)$ is analytic at $z = \infty$ (see Remark 1, p. 125) if $m < n$ or $m = n$, but not if $m > n$.

By the *order* N of a rational function $f(z)$, we mean the total number of zeros (or poles) of the function in the extended plane, counting each zero (or pole) a number of times equal to its order. It is easy to see that

$$N = \max(m, n),$$

i.e., N is the larger of the two integers m and n (or their common value if $m = n$). In fact, consider the three possibilities in turn:

1. If $m < n$, $f(z)$ has $k_1 + \cdots + k_p = m$ zeros at finite points and a zero of order $n - m$ at infinity, so that $f(z)$ has a total of

$$N = m + (n - m) = n = \max(m, n)$$

zeros. Moreover, $f(z)$ has $l_1 + \cdots + l_q = n$ poles at finite points and no pole at infinity, so that $f(z)$ also has a total of $N = n$ poles.

2. If $m = n$, $f(z)$ has neither a zero nor a pole at infinity, and precisely

$$m = k_1 + \cdots + k_p = l_1 + \cdots + l_q = m$$

zeros and poles at finite points.

3. If $m > n$, $f(z)$ has $l_1 + \cdots + l_q = n$ poles at finite points and a pole of order $m - n$ at infinity, so that $f(z)$ has a total of

$$N = n + (m - n) = m = \max(m, n)$$

poles. Moreover, $f(z)$ has $k_1 + \cdots + k_p = m$ zeros at finite points and no zero at infinity, so that $f(z)$ also has a total of $N = m$ zeros.

Now let A be an arbitrary complex number, and consider the roots of the equation

$$f(z) = \frac{P(z)}{Q(z)} = A, \tag{10.3}$$

or equivalently, the roots of the equation

$$F(z) = \frac{P(z) - AQ(z)}{Q(z)} = 0. \tag{10.4}$$

By the multiplicity of a root z_0 of equation (10.3), we mean the multiplicity of the same root z_0 of equation (10.4).

THEOREM 10.1. *Given a rational function $f(z)$ of order N and an arbitrary complex number A, the total number of roots of equation* (10.3) *is N (with due regard for multiplicity).*

Proof. If $A = 0$, or $A = \infty$, the result follows directly from the definition of the order of $f(z)$. Let r be the degree of the polynomial $P(z) - AQ(z)$ and s the order of the rational function $F(z)$. If $A \neq 0$, $A \neq \infty$, $m \neq n$, then $r = \max(m, n) = N$, and hence $s = \max(p, n) = N$. If $A \neq 0$, $A \neq \infty$, $m = n$, then $r \leqslant \max(m, n) = N$, and hence $s = \max(p, n) = N$.

Remark. As we know from Sec. 35, if $P_n(z)$ is a polynomial of degree n, the equation $P_n(z) = 0$ has exactly n roots (with due regard for multiplicity). Thus we see that this same property holds for rational functions, with the concept of *order* replacing that of *degree*.

Example. The rational function

$$f(z) = \frac{z^2 + 1}{z^2 - 1}$$

has two simple zeros $\pm i$ and two simple poles ± 1. If $A = 1$, the equation $f(z) = A$ becomes

$$\frac{z^2 + 1}{z^2 - 1} = 1$$

or

$$\frac{2}{z^2 - 1} = 0, \tag{10.5}$$

which obviously has no finite roots. However, (10.5) has a double root at ∞, since the degree of the denominator is two higher than that of the numerator.

It follows from Theorem 10.1 that a *rational function $w = f(z)$ of order $N > 0$ maps the extended plane onto itself in such a way that every point w has at most N distinct inverse images* (cf. the analogous property of

polynomials, given on p. 129). Moreover, only certain exceptional values of w can have fewer than N distinct inverse images [e.g., the point $w = 0$ if $f(z)$ has multiple zeros, or the point $w = \infty$ if $f(z)$ has multiple poles]. The following result is the natural generalization of Theorem 9.1:

THEOREM 10.2. *Let $f(z)$ be a rational function whose numerator $P(z)$ is of degree m and whose denominator $Q(z)$ is of degree n. Then there are at most $m + n$ points in the extended w-plane with fewer than $N = \max(m, n)$ distinct inverse images ($N > 1$).*

Proof. First suppose $A \neq 0$, $A \neq \infty$, and suppose that A has fewer than N distinct inverse images. Then the equation

$$f(z) = \frac{P(z)}{Q(z)} = A,$$

or equivalently

$$\frac{P(z) - AQ(z)}{Q(z)} = 0, \tag{10.6}$$

must have a multiple root. Any finite multiple root of (10.6) is a multiple root of the equation

$$P(z) - AQ(z) = 0, \tag{10.7}$$

and conversely. Moreover, any multiple root of (10.7) satisfies the equation

$$P'(z) - AQ'(z) = 0,$$

and hence also satisfies the equation

$$P'(z)Q(z) - P(z)Q'(z) = 0, \tag{10.8}$$

of degree no higher than $m + n - 1$. Equation (10.8) has no more than $m + n - 1$ distinct roots $\gamma_1, \ldots, \gamma_r$ ($1 \leqslant r \leqslant m + n - 1$). Since any finite multiple zero of $f(z)$ satisfies the equations

$$P(z) = 0, \qquad P'(z) = 0,$$

and since any finite multiple pole of $f(z)$ satisfies the equations

$$Q(z) = 0, \quad Q'(z) = 0,$$

all the finite multiple roots of (10.6) satisfy (10.8), and are therefore already included among the numbers $\gamma_1, \ldots, \gamma_r$, even if we let $A = 0$ or $A = \infty$. In other words, the numbers

$$f(\gamma_1), \ldots, f(\gamma_r), f(\infty) \qquad (1 \leqslant r \leqslant m + n - 1)$$

are the only values of A for which the equation $f(z) = A$ can have a multiple root, where we finally allow for the possibility that $z = \infty$ is a multiple root. Since at most $m + n$ of these numbers are distinct, the theorem is proved.[2]

[2] If $N = 0$ or $N = 1$, the theorem is meaningless.

Remark. The mapping $w = f(z)$ is conformal at all but a finite number of points, since the condition that z should not equal ∞, any of the numbers $\gamma_1, \ldots, \gamma_r$ or any of the zeros of $Q(z)$ is certainly sufficient to guarantee that the derivative

$$f'(z) = \frac{P'(z)Q(z) - P(z)Q'(z)}{[Q(z)]^2}$$

is finite and nonzero.[3]

44. The Group Property of Möbius Transformations

It follows from Theorems 10.1 and 10.2 that the rational function of order 1, i.e., the *fractional linear transformation* or *Möbius transformation*

$$w = L(z) = \frac{az + b}{cz + d} \qquad (ad - bc \neq 0)$$

is the only rational function which maps the extended plane onto itself in a one-to-one fashion. Here, we assume that the quantity $ad - bc$, called the *determinant* of the function $L(z)$, is nonzero, since otherwise $L(z)$ is identically equal to a constant mapping the extended plane into a single point. As we know from Sec. 34 (see also Problem 8.9), the mapping $w = L(z)$ is conformal at all points of the extended plane. This mapping (or its various special cases) plays an important role in a variety of problems encountered in the theory of functions of a complex variable, and it therefore merits a detailed study (to which we devote most of this chapter).

Let $\mathcal{M}$ denote the set of all Möbius transformations. Two such transformations

$$L_1(z) = \frac{a_1 z + b_1}{c_1 z + d_1}, \qquad L_2(z) = \frac{a_2 z + b_2}{c_2 z + d_2} \tag{10.9}$$

are regarded as *identical* if and only if $L_1(z) = L_2(z)$ for all z.

THEOREM 10.3. *A necessary and sufficient condition for the two Möbius transformations* (10.9) *to be identical is that*

$$a_2 = \lambda a_1, \quad b_2 = \lambda b_1, \quad c_2 = \lambda c_1, \quad d_2 = \lambda d_1 \qquad (\lambda \neq 0). \tag{10.10}$$

Proof. The sufficiency of the condition is obvious. To prove the necessity, suppose $L_1(z) = L_2(z)$. Then, in particular,

$$L_1(0) = L_2(0), \quad L_1(1) = L_2(1), \quad L_1(\infty) = L_2(\infty),$$

which means that[4]

$$\frac{b_1}{d_1} = \frac{b_2}{d_2} = p, \quad \frac{a_1 + b_1}{c_1 + d_1} = \frac{a_2 + b_2}{c_2 + d_2}, \quad \frac{a_1}{c_1} = \frac{a_2}{c_2} = q. \tag{10.11}$$

---

[3] However, the condition is not necessary. (For further details, see Problems 10.1 and 10.2.)

[4] Here we allow p or q to take the improper value ∞.

Substituting

$$b_1 = pd_1, \quad b_2 = pd_2, \quad a_1 = qc_1, \quad a_2 = qc_2$$

into the second of the equations (10.11), we obtain

$$\frac{qc_1 + pd_1}{c_1 + d_1} = \frac{qc_2 + pd_2}{c_2 + d_2}$$

or

$$(c_1d_2 - c_2d_1)(q - p) = 0.$$

But $q \neq p$, since otherwise

$$\frac{a_1}{c_1} = \frac{b_1}{d_1}, \qquad \text{i.e.,} \quad a_1d_1 - b_1c_1 = 0,$$

contrary to hypothesis, and therefore

$$\frac{c_1}{d_1} = \frac{c_2}{d_2}.$$

Together with (10.11), this implies (10.10), as required.

Remark. A Möbius transformation is not characterized by the value of its determinant, since the determinant is multiplied by λ^2 when the coefficients change as described by (10.10). It can only be asserted that the determinant remains nonzero under any substitution (10.10).

The transformation

$$U(z) = z,$$

which obviously belongs to the set $\mathscr{M}$, is called the *unit transformation* (or the *identity transformation*). By the inverse of a given transformation

$$w = \frac{az + b}{cz + d}, \tag{10.12}$$

we mean the transformation which assigns to each w its inverse image z under the transformation (10.12). Thus the transformation

$$z = \frac{dw - b}{-cw + a}$$

(whose coefficients are unique only to within a multiplicative constant, as in Theorem 10.3) is the inverse of the transformation (10.12). We denote the inverse of the transformation L by L^{-1}.

Given two arbitrary Möbius transformations

$$L_1(z) = \frac{a_1z + b_1}{c_1z + d_1}, \qquad L_2(z) = \frac{a_2z + b_2}{c_2z + d_2},$$

we define the *product* of L_1 and L_2 as the result of first carrying out one transformation and then carrying out the other. There are two possible

products corresponding to the two orders in which the transformations can be carried out. One product, written $L_1L_2(z)$, equals

$$\begin{aligned} L_1L_2(z) &= \frac{a_1[(a_2z + b_2)/(c_2z + d_2)] + b_1}{c_1[(a_2z + b_2)/(c_2z + d_2)] + d_1} \\ &= \frac{a_1(a_2z + b_2) + b_1(c_2z + d_2)}{c_1(a_2z + b_2) + d_1(c_2z + d_2)} \\ &= \frac{(a_1a_2 + b_1c_2)z + (a_1b_2 + b_1d_2)}{(c_1a_2 + d_1c_2)z + (c_1b_2 + d_1d_2)}, \end{aligned} \tag{10.13}$$

and the other, written $L_2L_1(z)$, is obtained from (10.13) by permuting the indices 1 and 2. Moreover, since

$$\begin{aligned} (a_1a_2 + b_1c_2)(c_1b_2 + d_1d_2) - (a_1b_2 + b_1d_2)(c_1a_2 + d_1c_2) \\ = (a_1d_1 - b_1c_1)(a_2d_2 - b_2c_2) \neq 0, \end{aligned}$$

each of the transformations $L_1L_2(z)$ and $L_2L_1(z)$ belongs to the set $\mathscr{M}$. In general, $L_1L_2(z) \neq L_2L_1(z)$, but obviously

$$LL^{-1}(z) = L^{-1}L(z) = U(z),$$

and

$$LU(z) = UL(z) = L(z)$$

for any $L(z) \in \mathscr{M}$.

Example. If

$$L_1(z) = \frac{z}{z + 1}, \qquad L_2(z) = \frac{z + 1}{z - 1},$$

then

$$L_1L_2(z) = \frac{z + 1}{2z}, \qquad L_2L_1(z) = -2z - 1.$$

Multiplication of transformations, as just defined, is an *associative* operation, i.e.,

$$(L_1L_2)L_3(z) = L_1(L_2L_3)(z). \tag{10.14}$$

To see this, we merely write $z_3 = L_3(z)$, and then both sides of (10.14) reduce at once to $L_1L_2(z_3)$. This associative property generalizes immediately to the product of an arbitrary number of transformations, and makes it unnecessary to use parentheses when writing products. For example, we have

$$L_1[L_2(L_3L_4)](z) = L_1L_2(L_3L_4)(z) = L_1(L_2L_3)L_4(z) = \cdots = L_1L_2L_3L_4(z).$$

Thus we have shown that the set $\mathscr{M}$ of Möbius transformations has the following properties:[5]

---

[5] Henceforth, for simplicity, we shall often omit the argument z, writing L_1 instead of $L_1(z)$, U instead of $U(z)$, etc.

1. $\mathscr{M}$ is *closed under multiplication*, i.e., if $L_1 \in \mathscr{M}$, $L_2 \in \mathscr{M}$, then $L_1L_2 \in \mathscr{M}$, $L_2L_1 \in \mathscr{M}$.
2. Multiplication is associative.
3. There is an element $U \in \mathscr{M}$ such that $LU = UL = L$ for any $L \in \mathscr{M}$.
4. For each $L \in \mathscr{M}$, there is an element $L^{-1} \in \mathscr{M}$ such that $LL^{-1} = L^{-1}L = U$.

In algebraic language, these four properties are summarized by saying that $\mathscr{M}$ is a group of transformations.[6]

45. The Circle-Preserving Property of Möbius Transformations

We now prove that any Möbius transformation carries a straight line or a circle into another straight line or circle. We call this the *circle-preserving property*, since a straight line can be regarded as a limiting case of a circle (corresponding to infinite radius). The entire linear transformation $L(z) = \alpha z + \beta$ ($\alpha \neq 0$) is obviously circle-preserving, since the mapping $w = L(z)$ is just a shift (if $\alpha = 1$), or a shift combined with a rotation and a uniform magnification (if $\alpha \neq 1$), as discussed in Sec. 33.

LEMMA. *The transformation*

$$w = \Lambda(z) = \frac{1}{z} \tag{10.15}$$

is circle-preserving.

Proof. The equation of any straight line or circle in the z-plane can be written in the form

$$A(x^2 + y^2) + 2Bx + 2Cy + D = 0, \tag{10.16}$$

where we have a straight line if $A = 0$ and at least one of the numbers B, C is nonzero, and a circle if $A \neq 0$ and $B^2 + C^2 - AD > 0$. Since

$$x^2 + y^2 = z\bar{z}, \quad 2x = z + \bar{z}, \quad 2y = -i(z - \bar{z}),$$

where $\bar{z} = x - iy$ is the complex conjugate of $z = x + iy$, we can rewrite (10.16) as

$$Az\bar{z} + \bar{E}z + E\bar{z} + D = 0, \tag{10.17}$$

where $E = B + iC$. It is easy to see that equation (10.17), where A and D are real and E is complex, is the equation of a straight line if and only if $A = 0$, $E \neq 0$, and the equation of a circle if and only if $A \neq 0$, $E\bar{E} - AD > 0$.

---

[6] See e.g., G. Birkhoff and S. MacLane, *op. cit.*, Chap, 6, Sec. 2, or V. I. Smirnov, *Linear Algebra and Group Theory* (translated by R. A. Silverman), McGraw-Hill Book Co., New York (1961), Sec. 62.

We now find the image of the curve with equation (10.17) under the transformation (10.15). Replacing z by $1/w$ in (10.17), we obtain

$$A\frac{1}{w\bar{w}} + \bar{E}\frac{1}{w} + E\frac{1}{\bar{w}} + D = 0$$

or

$$Dw\bar{w} + Ew + \bar{E}\bar{w} + A = 0. \tag{10.18}$$

Equation (10.18) has the same form as equation (10.17), with D, $\bar{E}$ and A substituted for A, E and D, respectively. It follows that (10.18) is the equation of a straight line if $D = 0$, since then either $A = 0$ and $E \neq 0$ if (10.17) is the equation of a straight line, or else $A \neq 0$ and $E\bar{E} - AD = E\bar{E} > 0$ (so that $E \neq 0$ again) if (10.17) is the equation of a circle. Moreover, (10.18) is the equation of a circle if $D \neq 0$, since then either $A \neq 0$ and $E\bar{E} - AD > 0$ if (10.17) is the equation of a circle, or else $A = 0$ and $E \neq 0$ (so that $E\bar{E} - AD = E\bar{E} > 0$ again) if (10.17) is the equation of a straight line.

THEOREM 10.4. *Every Möbius transformation*

$$w = L(z) = \frac{az + b}{cz + d} \qquad (ad - bc \neq 0) \tag{10.19}$$

is circle-preserving.

Proof. If $c = 0$, (10.19) reduces to an entire linear transformation and hence is circle-preserving. If $c \neq 0$, (10.19) can be written in the form

$$w = \frac{a}{c} + \frac{bc - ad}{c(cz + d)}.$$

Setting

$$z_1 = L_1(z) = cz + d, \quad z_2 = \Lambda(z_1) = \frac{1}{z_1},$$
$$w = L_2(z_2) = \frac{a}{c} + \frac{bc - ad}{c} z_2,$$

we can write $L(z)$ as a product

$$L = L_2 \Lambda L_1$$

of three transformations which are all circle-preserving (use the lemma). It follows that L itself is circle-preserving.

COROLLARY. *Let $\delta = -d/c$ be the pole of the function* (10.19). *Then* (10.19) *transforms every straight line or circle which passes through δ into a straight line, and every other straight line or circle into a circle.*

Proof. If the circle or straight line passes through δ, its image under (10.19) contains the point at infinity, and hence must be a straight line,

since it cannot be a circle (no circle contains ∞). Similarly, if the circle or straight line does not pass through δ, its image does not contain the point at infinity, and hence must be a circle, since it cannot be a straight line (every straight line contains ∞).

Remark. Let $w = L(z)$ be any Möbius transformation, let γ be a straight line or circle in the z-plane, and let $\Gamma = L(\gamma)$ be the image of γ in the w-plane (Γ is itself a straight line or a circle). The two domains G_1 and G_2 with boundary γ are either two half-planes or the interior and exterior of a circle. Let $L(G_1)$ and $L(G_2)$ be the images of these two domains under the mapping $w = L(z)$. We now show that $L(G_1)$ and $L(G_2)$ are the two domains whose common boundary is the curve Γ.[7]

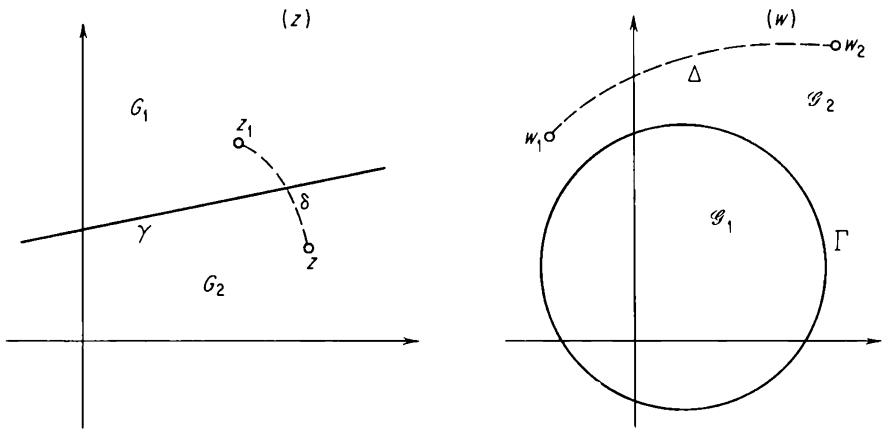

FIGURE 10.1

First suppose $z_1 \in G_1$, $z_2 \in G_2$, and let $w_1 = L(z_2)$, $w_2 = L(z_2)$. Then $w_1 \notin \Gamma$, $w_2 \notin \Gamma$, since $z_1 \notin \gamma$, $z_2 \notin \gamma$, and hence w_1 and w_2 must belong to the union of the two (disjoint) domains into which Γ divides the extended w-plane. If w_1 and w_2 both belong to one of these two domains, we can join w_1 to w_2 by a line segment or circular arc Δ which does not intersect Γ (see Figure 10.1). The inverse image of Δ in the z-plane must be a line segment or circular arc δ, which joins z_1 to z_2 and does not intersect γ. But the existence of δ contradicts the assumption that z_1 and z_2 belong to different domains G_1 and G_2. Therefore, if z_1 and z_2 belong to different domains with boundary γ, their images w_1 and w_2 must belong to different domains with boundary Γ.

We now denote the domains containing w_1 and w_2 by $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively. If z is an arbitrary point in G_1, then, since z and z_2 belong to different domains G_1 and G_2, their images w and w_2 belong to different

[7] Of course, this result follows at once from Theorem 6.3, knowledge of which is not presupposed here.

domains $\mathcal{G}_1$ and $\mathcal{G}_2$. But $w_2 \in \mathcal{G}_2$, and hence $w \in \mathcal{G}_1$, i.e., $L(z) \in \mathcal{G}_1$, if $z \in G_1$. Similarly, $L(z) \in \mathcal{G}_2$ if $z \in G_2$, and hence

$$\mathcal{G}_1 \supset L(G_1), \qquad \mathcal{G}_2 \supset L(G_2). \tag{10.20}$$

Conversely, let w be an arbitrary point in $\mathcal{G}_1$. Then w must be the image of a point z in G_1 or G_2. But $z \in G_2$ implies $w \in \mathcal{G}_2$, contrary to hypothesis, and hence $z \in G_1$, i.e., $\mathcal{G}_1 \subset L(G_1)$. Similarly, we find that $\mathcal{G}_2 \subset L(G_2)$. It follows by comparison with (10.20) that

$$\mathcal{G}_1 = L(G_1), \qquad \mathcal{G}_2 = L(G_2),$$

i.e., the two domains with boundary Γ are just the images of the two domains G_1 and G_2, as asserted. Moreover, to determine which of the two domains with boundary Γ is actually the image of a given domain G_1 with boundary γ, it is sufficient to locate the image w_1 of any point $z_1 \in G_1$, for then the domain $\mathcal{G}_1$ containing w_1 is the image of G_1.

46. Fixed Points of a Möbius Transformation. Invariance of the Cross Ratio

By a *fixed point* of a transformation or mapping $w = f(z)$, we mean a point which is carried into itself by the transformation. Obviously, every such point is a solution of the equation

$$z = f(z).$$

Moreover, every point of the z-plane is trivially a fixed point of the unit transformation $U(z) = z$.

THEOREM 10.5. *Every Möbius transformation different from the unit transformation has two fixed points, which in certain cases coalesce into a single fixed point.*

Proof. First let $c = 0$ $(d \neq 0)$, so that $L(z)$ reduces to the entire linear transformation

$$L(z) = \alpha z + \beta \qquad \left(\alpha = \frac{a}{d}, \quad \beta = \frac{b}{d}\right).$$

Then, since $L(\infty) = \infty$, one fixed point is the point at infinity. If $\alpha \neq 1$, there exists another fixed point determined from the equation

$$z = \alpha z + \beta,$$

i.e., the point $\beta/(1 - \alpha)$, but if $\alpha = 1$, $\beta \neq 0$, there is no finite fixed point. Moreover, if $\alpha \neq 1$, $\beta \neq 0$, the finite fixed point $\beta/(1 - \alpha)$ approaches ∞ as $\alpha \to 1$. Therefore, in the case of the transformation

$$L(z) = z + \beta \qquad (\beta \neq 0),$$

the point at infinity can be regarded as two fixed points which have coalesced.

Now let $c \neq 0$, so that

$$L(\infty) = \frac{a}{c} \neq \infty,$$

i.e., the point at infinity is not a fixed point.[8] Similarly, the pole $\delta = -d/c$ of the transformation is not a fixed point, since

$$L(\delta) = \infty \neq \delta.$$

Assuming that $z \neq \infty$ and $z \neq \delta$, we solve the equation

$$z = \frac{az + b}{cz + d}$$

or

$$cz^2 - (a - d)z - b = 0,$$

obtaining

$$z = \frac{a - d \pm \sqrt{(a - d)^2 + 4bc}}{2c}.$$

If $(a - d)^2 + 4bc \neq 0$, we obtain two different finite fixed points; if $(a - d)^2 + 4bc = 0$, these two points coalesce to form a single finite fixed point $(a - d)/2c$.

COROLLARY. *The only Möbius transformation with more than two fixed points is the unit transformation $U(z) = z$, for which all points are fixed points.*

THEOREM 10.6. *A sufficient condition for two Möbius transformations $L(z)$ and $\Lambda(z)$ to be identical is that the equation*

$$L(z) = \Lambda(z)$$

hold for three distinct points z_1, z_2 and z_3. In particular, there cannot exist two distinct Möbius transformations taking three given values w_1, w_2, w_3 at three given distinct points z_1, z_2, z_3.

Proof. It follows from

$$L(z_k) = \Lambda(z_k) = w_k \qquad (k = 1, 2, 3)$$

that

$$\Lambda^{-1}(w_k) = z_k \qquad (k = 1, 2, 3),$$

and hence

$$\Lambda^{-1}L(z_k) = z_k \qquad (k = 1, 2, 3).$$

[8] Therefore, in the class of Möbius transformations, an entire linear transformation is characterized by the fact that at least one of its fixed points is the point at infinity.

Therefore the transformation $\Lambda^{-1}L$ has three distinct fixed points, which, according to the corollary, implies

$$\Lambda^{-1}L = U,$$

where U is the unit transformation. Multiplying both sides of (10.21) by Λ from the left, we obtain

$$\Lambda(\Lambda^{-1}L) = \Lambda U,$$

which implies

$$L = \Lambda,$$

as asserted.

We now set about determining the (unique) Möbius transformation $w = L(z)$ carrying the points z_1, z_2, z_3 into the points w_1, w_2, w_3. First we consider the problem of finding the special Möbius transformation $w = \Lambda(z)$ carrying three finite points z_1, z_2, z_3 into the points 0, ∞, 1. Since the function

$$w = \Lambda(z) = \frac{az + b}{cz + d}$$

vanishes for $z = z_1$ and becomes infinite for $z = z_2$ if and only if z_1 is a zero of the numerator and z_2 is a zero of the denominator, it follows that

$$w = \Lambda(z) = \frac{a}{c}\frac{z - z_1}{z - z_2}.$$

But w must equal 1 for $z = z_3$, i.e.,

$$1 = \frac{a}{c}\frac{z_3 - z_1}{z_3 - z_2},$$

which implies[9]

$$\frac{a}{c} = 1 : \frac{z_3 - z_1}{z_3 - z_2}.$$

Therefore the Möbius transformation carrying z_1, z_2, z_3 into 0, ∞, 1 is

$$w = \Lambda(z) = \frac{z - z_1}{z - z_2} : \frac{z_3 - z_1}{z_3 - z_2}. \tag{10.21}$$

Next we determine the more general Möbius transformation $w = L(z)$ satisfying

$$L(z_1) = w_1, \quad L(z_2) = w_2, \quad L(z_3) = w_3,$$

where w_1, w_2, w_3 are three arbitrary (but distinct) finite points. As we have just seen, the transformation

$$\Lambda_1(z) = \frac{z - w_1}{z - w_2} : \frac{w_3 - w_1}{w_3 - w_2} \tag{10.22}$$

---

[9] By $x:y$ is meant the *ratio of x to y*, i.e., the quantity x/y.

carries the points w_1, w_2, w_3 into the points 0, ∞, 1. Therefore the transformation $\Lambda_1 L$ carries the points z_1, z_2, z_3 into the points 0, ∞, 1, so that

$$\Lambda_1 L(z) = \Lambda(z) = \frac{z - z_1}{z - z_2} : \frac{z_3 - z_1}{z_3 - z_2}. \tag{10.23}$$

Multiplying both sides of the equation

$$\Lambda_1 L(z) = \Lambda(z) \tag{10.24}$$

by Λ_1^{-1} from the left, we obtain

$$L(z) = \Lambda_1^{-1}\Lambda(z).$$

This solves the problem, since the functions $\Lambda(z)$ and $\Lambda_1(z)$, and hence $\Lambda_1^{-1}(z)$, are known [cf. (10.23) and (10.22)]. However, it is more convenient to use (10.24) directly, after writing $w = L(z)$. The result is

$$\Lambda_1(w) = \Lambda(z)$$

or

$$\frac{w - w_1}{w - w_2} : \frac{w_3 - w_1}{w_3 - w_2} = \frac{z - z_1}{z - z_2} : \frac{z_3 - z_1}{z_3 - z_2}, \tag{10.25}$$

which expresses the Möbius transformation $w = L(z)$ in implicit form.

Remark. In finding the Möbius transformation carrying the points z_1, z_2, z_3 into the points w_1, w_2, w_3, it was assumed that all six points are finite. However, the case of infinite points is easily handled. For example, the transformation carrying the points ∞, z_2, z_3 into the points 0, ∞, 1 has the form

$$w = \Lambda(z) = \frac{1}{z - z_2} : \frac{1}{z_3 - z_2},$$

which can be found by inspection or by writing (10.21) in the form

$$w = \frac{\dfrac{z}{z_1} - 1}{z - z_2} : \frac{\dfrac{z_3}{z_1} - 1}{z_3 - z_2}$$

and then taking the limit as $z_1 \to \infty$. Therefore (10.25) is replaced by

$$\frac{w - w_1}{w - w_2} : \frac{w_3 - w_1}{w_3 - w_2} = \frac{1}{z_1 - z_2} : \frac{1}{z_3 - z_2},$$

where it is assumed that the points w_1, w_2, w_3 are finite. Similarly, the transformation carrying the points z_1, ∞, z_3 into the points 0, ∞, 1 has the form

$$\Lambda(z) = (z - z_1) : (z_3 - z_1),$$

and hence we have

$$\frac{w - w_1}{w - w_2} : \frac{w_3 - w_1}{w_3 - w_2} = (z - z_1) : (z_3 - z_1),$$

instead of (10.25). Finally, the transformation carrying the points z_1, z_2, ∞ into the points 0, ∞, 1 has the form

$$\Lambda(z) = \frac{z - z_1}{z - z_2},$$

and hence we have

$$\frac{w - w_1}{w - w_2} : \frac{w_3 - w_1}{w_3 - w_2} = \frac{z - z_1}{z - z_2},$$

instead of (10.25).

In just the same way, we have to replace the left-hand side of (10.25) by

$$\frac{1}{w - w_2} : \frac{1}{w_3 - w_2}, \quad (w - w_1) : (w_3 - w_1) \quad \text{or} \quad \frac{w - w_1}{w - w_2},$$

depending on whether $w_1 = \infty$, $w_2 = \infty$ or $w_3 = \infty$. As a result, we arrive at the following mnemonic rule: If $z_k = \infty$ or $w_l = \infty$ ($k, l = 1, 2, 3$), the differences involving z_k or w_l have to be replaced by 1. The reader can easily verify this rule by taking the appropriate limits (as $z_k \to \infty$ or $w_l \to \infty$) in equation (10.25).

Equation (10.25) implies an important general property of Möbius transformations. If a, b, c and d are arbitrary distinct finite complex numbers, then the ratio

$$\frac{c - a}{c - b} : \frac{d - a}{d - b},$$

denoted by (a, b, c, d), is called the *cross ratio* (or *anharmonic ratio*) of the four numbers (or points) a, b, c and d. If one of the four points a, b, c, d is the point at infinity, we define the cross ratio as the limit of the cross ratio of four finite points, three of which coincide with the three given finite points, as the fourth point approaches infinity. Thus, according to this definition, we have

$$\begin{aligned}
(\infty, b, c, d) &= \frac{1}{c - b} : \frac{1}{d - b}, \\
(a, \infty, c, d) &= (c - a) : (d - a), \\
(a, b, \infty, d) &= 1 : \frac{d - a}{d - b}, \\
(a, b, c, \infty) &= \frac{c - a}{c - b}.
\end{aligned}$$

Now let $w = L(z)$ be an arbitrary Möbius transformation, and let A, B, C, D be the points into which $L(z)$ maps four arbitrary (but distinct) points a, b, c, d. Since the points A, B and D are the images of the points a, b and d, it follows from (10.25) that the relation between z and $w = L(z)$ is given by

$$\frac{w - A}{w - B} : \frac{D - A}{D - B} = \frac{z - a}{z - b} : \frac{d - a}{d - b},$$

where differences involving the point at infinity have to be replaced by 1. Moreover, since the point C is the image of the point c, we have

$$\frac{C-A}{C-B} : \frac{D-A}{D-B} = \frac{c-a}{c-b} : \frac{d-a}{d-b},$$

(where again differences involving the point at infinity have to be replaced by 1), or equivalently

$$(A, B, C, D) = (a, b, c, d).$$

In other words, *the cross ratio of any four distinct points is invariant under a Möbius transformation.*

47. Mapping of a Circle onto a Circle[10]

Using the circle-preserving property of Möbius transformations and the possibility of mapping any given triple of distinct points z_1, z_2, z_3 into any other given triple of distinct points w_1, w_2, w_3, we obtain the following basic result:

THEOREM 10.7. *Let γ and Γ be any two straight lines or circles, and let z_1, z_2, z_3 and w_1, w_2, w_3 be any two triples of distinct points belonging to γ and Γ, respectively. Then there exists a Möbius transformation $w = L(z)$ mapping γ onto Γ in such a way that*

$$L(z_k) = w_k \qquad (k = 1, 2, 3). \tag{10.26}$$

Proof. Construct the Möbius transformation $w = L(z)$ satisfying the conditions (10.26), which according to Theorem 10.6 and the subsequent construction, exists and is unique. According to Theorem 10.4, $w = L(z)$ maps the straight line or circle γ onto another straight line or circle Γ^*. But since γ goes through the points z_1, z_2 and z_3, Γ^* must go through the points w_1, w_2 and w_3. Moreover, since two different straight lines or circles cannot be drawn through the same three points, Γ^* must coincide with Γ, as asserted.

Remark. Again consider two arbitrary straight lines or circles γ and Γ (which may coincide). Let G be one of the two domains with boundary γ, and let $\mathscr{G}$ be one of the two domains with boundary Γ, so that G is either a half-plane, the interior of a circle or the exterior of a circle, and the same is true of $\mathscr{G}$. We now show how to map G onto $\mathscr{G}$. Choose an arbitrary triple of distinct points z_1, z_2, z_3 on γ, and suppose an observer moving along γ in the direction from z_1 to z_3 through z_2 finds the domain G on his *left*, say. Next choose a triple of distinct points w_1, w_2, w_3 on Γ such that an observer moving along Γ in the direction from w_1 to w_3 through w_2 finds the domain $\mathscr{G}$ on his left, but let w_1, w_2, w_3 be otherwise arbitrary. As in Theorem

[10] As usual, a straight line is regarded as a limiting case of a circle (cf. p. 168).

10.7, we form the Möbius transformation $w = L(z)$ which satisfies the conditions (10.26) and hence maps γ onto Γ. Then $w = L(z)$ also maps G onto $\mathcal{G}$, i.e., $\mathcal{G} = L(G)$. In fact, if δ is a segment of the normal to the curve γ drawn from the point z_2 and pointing into the interior of G, so that an observer at z_2 facing in the direction established on γ finds δ on his left, then, since the mapping $w = L(z)$ is conformal, an observer at w_2 facing in the direction established on Γ will also find the image $\Delta = L(\delta)$, which is a line segment or circular arc, on his left (see Figure 10.2 and Remark 2, p. 156). Therefore $\Delta \subset \mathcal{G}$, and hence $\mathcal{G}$ contains images of certain points belonging to G (i.e., the points of the segment δ). But, according to the remark on p. 170, $L(G)$ is one of the two domains with boundary $\Gamma = L(\gamma)$, in fact just the domain containing the image of any point in G. In other words, $\mathcal{G} = L(G)$, as asserted.

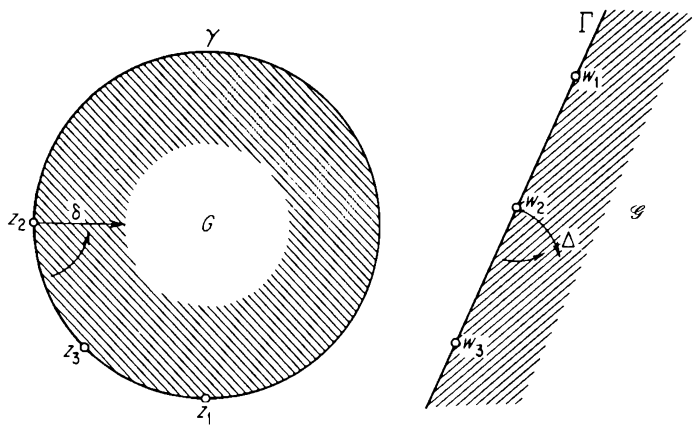

FIGURE 10.2

Example. *Find a conformal mapping of the upper half-plane* $\operatorname{Im} z > 0$ *onto the interior of the unit circle.*

To solve this problem, we choose $z_1 = -1$, $z_2 = 0$, $z_3 = 1$, say, so that the upper half-plane is on the left of an observer moving along the real axis in the direction from z_1 to z_3 through z_2. We also choose three points w_1, w_2, w_3 on the unit circle, such that the interior of the circle is on the left of an observer moving along the circle in the direction from w_1 to w_3 through w_2. For simplicity, we choose $w_1 = 1$, $w_2 = i$, $w_3 = -1$. Then the desired Möbius transformation satisfies the conditions $L(z_k) = w_k$, $k = 1, 2, 3$, and can be represented in the form

$$\frac{w-1}{w-i} : \frac{-1-1}{-1-i} = \frac{z+1}{z} : \frac{1+1}{1}$$

or

$$w = \frac{z-i}{iz-1},$$

where we have used (10.25).

48. Symmetry Transformations

Let z_1 and z_2 be two points which are symmetric with respect to a given straight line γ, i.e., such that γ is the perpendicular bisector of the line segment joining z_1 and z_2. By definition, the straight line passing through z_1 and z_2 is orthogonal to γ. Moreover, the center of any circle δ passing through z_1 and z_2 lies on γ, and hence δ is also orthogonal to γ. It is easy to see that the converse is true as well, i.e., if every straight line or circle passing through a pair of points z_1 and z_2 is orthogonal to a given straight line γ, then z_1 and z_2 are symmetric with respect to γ. Generalizing the concept of symmetry with respect to a straight line, we introduce the following definition: *Two points z_1 and z_2 are symmetric with respect to a given circle γ if and only if every straight line or circle passing through z_1 and z_2 is orthogonal to γ.*

THEOREM 10.8. *Let z_1 and z_2 be any two points symmetric with respect to a given straight line or circle γ, and let $w = L(z)$ be any Möbius transformation. Then the points $w_1 = L(z_1)$ and $w_2 = L(z_2)$ are symmetric with respect to the straight line or circle $\Gamma = L(\gamma)$.*[11]

Proof. We have to show that an arbitrary straight line or circle Δ passing through w_1 and w_2 is orthogonal to Γ. Let $z = L^{-1}(w)$ be the inverse of the transformation $w = L(z)$. Clearly, L^{-1} is also a Möbius transformation, and

$$L^{-1}(w_1) = z_1, \quad L^{-1}(w_2) = z_2, \quad L^{-1}(\Gamma) = \gamma.$$

Moreover, $\delta = L^{-1}(\Delta)$ is a straight line or circle passing through z_1 and z_2. Since z_1 and z_2 are symmetric with respect to γ, by hypothesis, it follows that δ is orthogonal to γ. But then, since the mapping $w = L(z)$ is conformal (see Sec. 33), $\Delta = L(\delta)$ is orthogonal to Γ, and the proof is complete.

COROLLARY. *There is only one point z_2 symmetric to a given point z_1 with respect to a given straight line or circle γ.*

Proof. If γ is a straight line, the statement is obvious. Thus let γ be a circle, and suppose that besides z_2, there is another point $z_2' \neq z_2$ symmetric to z_1 with respect to γ. Choosing a Möbius transformation $w = L(z)$ mapping γ onto a straight line Γ, we find that $w_2 = L(z_2)$ and $w_2' = L(z_2')$ are two distinct points symmetric with respect to Γ, which is impossible.

Remark. Suppose $w = L(z)$ maps a straight line or circle γ onto a circle Γ with center w_1, and let z_1 be the inverse image of w_1. Then the point z_2

[11] It is in this sense that Möbius transformations are said to be *symmetry-preserving*.

symmetric to z_1 with respect to γ must be mapped into the point at infinity. To see this, we note that $w_2 = \infty$ is symmetric to $w_1 = 0$ with respect to the circle Γ, since any straight line or circle passing through 0 and ∞, i.e., any straight line passing through the center of Γ, is orthogonal to Γ. The uniqueness of w_2 follows from the corollary.

Let γ be an arbitrary straight line or circle. A transformation of the extended plane into itself, which carries each point z into the point z^* symmetric to z with respect to γ is called a *symmetry transformation* with respect to γ or a *reflection* in γ. In the case where γ is a circle, the transformation is also called an *inversion* in γ. We now derive analytical expressions for symmetry transformations.

First let γ be a straight line with an assigned direction, and consider reflection in γ. The straight line γ is completely characterized by one of its points a and by the unit vector

$$e^{i\theta} = \cos\theta + i\sin\theta$$

pointing in the direction of γ. Suppose we carry out the entire linear transformation

$$z = L(w) = a + e^{i\theta}w, \tag{10.27}$$

which obviously maps the real axis onto γ, since (10.27) corresponds to a shift by the vector a (carrying the origin of coordinates to the point a), followed by a rotation through the angle θ about the point a. Since the inverse transformation $w = L^{-1}(z)$ maps γ onto the real axis, it maps every pair of points z and z^* symmetric with respect to γ into a pair of points w and w^* symmetric with respect to the real axis. But the points w and w^* are represented by two conjugate complex numbers, i.e.,

$$w = t, \qquad w^* = \bar{t}. \tag{10.28}$$

Therefore $z - a = e^{i\theta}t$, and

$$\overline{z - a} = e^{-i\theta}\bar{t}, \qquad z^* - a = e^{i\theta}w^* = e^{i\theta}\bar{t}. \tag{10.29}$$

Eliminating $\bar{t}$ from (10.29), we obtain

$$z^* - a = e^{2i\theta}\overline{(z - a)}. \tag{10.30}$$

According to (10.30), reflection in a straight line γ going through a point a at an angle θ with the real axis can be accomplished by first constructing the vector $\overline{z - a}$ which is the reflection of the vector $z - a$ in the real axis, and then rotating $\overline{z - a}$ through the angle 2θ about the point a.

Next let γ be a circle, and consider inversion in γ. Let R $(0 < R < \infty)$ be the radius and a the center of γ. We begin by finding a Möbius transformation

which maps γ onto the real axis. The simplest approach is to choose the transformation

$$z = L(w) = a + R\frac{1 + iw}{1 - iw},$$

which maps the three points $w_1 = -1$, $w_2 = 0$, $w_3 = 1$ of the real axis into the points $z_1 = a - iR$, $z_2 = a + R$, $z_3 = a + iR$ of the circle γ. The inverse transformation $w = L^{-1}(z)$ maps γ onto the real axis, and maps every pair of points z and z^* symmetric with respect to γ into a pair of points w and w^* symmetric with respect to the real axis. As before, the points w and w^* are represented by two conjugate complex numbers (10.28). Therefore

$$z - a = R\frac{1 + it}{1 - it},$$

and

$$\overline{z - a} = R\frac{1 - i\bar{t}}{1 + i\bar{t}}, \qquad z^* - a = R\frac{1 + i\bar{t}}{1 - i\bar{t}}. \tag{10.31}$$

Multiplying the two equations (10.31), we obtain

$$\overline{(z - a)}(z^* - a) = R^2$$

or

$$z^* - a = \frac{R^2}{z - a}. \tag{10.32}$$

In particular, it follows from (10.32) that

$$\operatorname{Arg}(z^* - a) = -\operatorname{Arg}\overline{(z - a)} = \operatorname{Arg}(z - a)$$

and

$$|z - a|\,|z^* - a| = R^2.$$

Therefore the points z and z^* lie on the same ray emanating from the center of γ, and the product of their distances from the center of γ equals the square of the radius of γ. These two conditions, equivalent to formula (10.32), determine the position of one of the points z, z^* with respect to the other, and completely characterize the operation of inversion in the circle γ, with equation $|z - a| = R$.

Remark. We note that any symmetry transformation reduces to consecutive application of a linear transformation (entire or fractional), followed by reflection in the real axis. In fact, according to (10.30), reflection in a straight line can be represented in the form

$$z_1 = \bar{a} + e^{-2i\theta}(z - a), \qquad z^* = \bar{z}_1,$$

while, according to (10.32), reflection in a circle can be represented in the form

$$z_1 = \bar{a} + \frac{R^2}{z - a}, \qquad z^* = \bar{z}_1.$$

Since any Möbius transformation is conformal and circle-preserving, and since reflection in the real axis has the same properties, except that while preserving the magnitudes of angles it reverses the directions in which they are measured, we see that the most general symmetry transformation is a conformal mapping of the second kind (see p. 120) which is also circle-preserving.

49. Examples

We now give two examples illustrating the use of the symmetry-preserving property of Möbius transformations.

Example 1. *Find a conformal mapping of the upper half-plane* Π_U: $\operatorname{Im} z > 0$ *onto the disk* K: $|w| < R$, *such that a given point* $\alpha \in \Pi_U$ *is mapped into the center of* K.[12]

Any such mapping

$$w = L(z) = \frac{az + b}{cz + d},$$

provided it exists, vanishes for $z = \alpha$, and hence α is the zero of $L(z)$. But the point $\bar{\alpha}$ symmetric to α with respect to the real axis must be mapped into the point symmetric to the center $w = 0$ of K with respect to the boundary of K, i.e., the circle C: $|w| = R$. Therefore $\bar{\alpha}$ must be mapped into the point at infinity (see the remark on p. 178), so that $L(\bar{\alpha}) = \infty$ and $\bar{\alpha}$ is the pole of $L(z)$. It follows that $L(z)$ has the form

$$w = L(z) = \frac{a(z - \alpha)}{c(z - \bar{\alpha})} = \lambda \frac{z - \alpha}{z - \bar{\alpha}}, \tag{10.33}$$

where λ is a nonzero complex number.

We now show that (10.33) maps Π_U onto K, with $z = \alpha$ going into $w = 0$, if we choose $|\lambda| = R$. Since $L(\alpha) = 0$ for any λ, by construction, it is sufficient to show that (10.33) maps the real axis onto the circle C. If $z = x$ is an arbitrary real number, then $x - \alpha$ and $x - \bar{\alpha}$ are complex conjugates, and hence

$$|w| = |L(x)| = \left|\lambda \frac{x - \alpha}{x - \bar{\alpha}}\right| = |\lambda| \left|\frac{x - \alpha}{x - \bar{\alpha}}\right| = |\lambda| = R.$$

Therefore (10.33) maps the real axis *into* C. But since any three distinct points of the real axis are mapped into three distinct points of C, it follows from Theorem 10.7 that the real axis is mapped *onto* C.

In (10.33) the argument of λ is left unspecified. The geometric reason for this indeterminacy is clear: Going from one value of λ to another in (10.33),

---

[12] In expressions like Π_U: $\operatorname{Im} z > 0$ and K: $|w| < R$, the colon means "with equation."

while keeping $|\lambda| = R$ fixed, is equivalent to changing the arguments of all points w by the same quantity, i.e., to rotating the disk K about its center $w = 0$. Such a rotation transforms K into itself while leaving its center fixed, and hence does not violate the conditions of the problem. Thus, if the problem is to have a unique solution, we must impose an extra condition on $L(z)$. For example, we might require either that

1. *A given point $x = x_0$ of the real axis should go into the point $w = R$ of the circle C*, or that
2. *The derivative $L'(\alpha)$ should be a positive real number.* (Geometrically, this means that the mapping does not change the slopes of tangents to curves passing through the point α.)

Imposing condition 1, we find from (10.33) that

$$R = L(x_0) = \lambda \frac{x_0 - \alpha}{x_0 - \bar{\alpha}},$$

so that

$$\lambda = R \frac{x_0 - \bar{\alpha}}{x_0 - \alpha},$$

and hence

$$L(z) = R \frac{x_0 - \bar{\alpha}}{x_0 - \alpha} \frac{z - \alpha}{z - \bar{\alpha}}.$$

Moreover, we still have

$$|\lambda| = R \left| \frac{x_0 - \bar{\alpha}}{x_0 - \alpha} \right| = R,$$

as required. Imposing condition 2, with $\alpha = \xi + i\eta$ ($\eta > 0$), we have

$$L'(\alpha) = \frac{\lambda}{\alpha - \bar{\alpha}} = \frac{\lambda}{2\eta i},$$

so that λ/i is a positive real number. But, on the other hand, $|\lambda|$ must equal R. Therefore $\lambda = iR$, and

$$L(z) = iR \frac{z - \alpha}{z - \bar{\alpha}}.$$

Example 2. *Find a conformal mapping of the disk K: $|z| < R$ onto itself such that a given point $\alpha \in K$ is mapped into the center of K.*

Any such mapping

$$w = L(z) = \frac{az + b}{cz + d},$$

provided it exists, vanishes for $z = \alpha$, and hence α is the zero of $L(z)$. But the point α^* symmetric to α with respect to the boundary of K, i.e., the circle C: $|z| = R$, must be mapped into a point symmetric to the center $w = 0$ of K

with respect to C. Therefore α^* must be mapped into the point at infinity, so that $L(\alpha^*) = \infty$ and α^* is the pole of $L(z)$. It follows that $L(z)$ has the form

$$w = L(z) = \lambda \frac{z - \alpha}{z - \alpha^*} \tag{10.34}$$

[cf. (10.33)], where λ is a nonzero complex number. According to (10.32), the point α^* symmetric to α with respect to C is

$$\alpha^* = \frac{R^2}{\bar{\alpha}},$$

and hence (10.34) becomes

$$w = L(z) = -\lambda\bar{\alpha} \frac{z - \alpha}{R^2 - \bar{\alpha}z} = \mu \frac{z - \alpha}{R^2 - \bar{\alpha}z}. \tag{10.35}$$

We now show that (10.35) maps K onto K, with $z = \alpha$ going into $w = 0$, if we choose $|\mu| = R^2$. Since $L(\alpha) = 0$ for any μ, by construction, it is sufficient to show that (10.35) maps the circle C: $|z| = R$ onto itself. If

$$z = Re^{i\theta} \qquad (0 \leqslant 0 < 2\pi)$$

is an arbitrary point of C, then

$$w = L(Re^{i\theta}) = \mu \frac{re^{i\theta} - \alpha}{R^2 - \bar{\alpha}\, Re^{i\theta}} = \frac{\mu}{Re^{i\theta}} \frac{Re^{i\theta} - \alpha}{Re^{-i\theta} - \bar{\alpha}},$$

and hence

$$|w| = |L(Re^{i\theta})| = \left| \frac{\mu}{Re^{i\theta}} \frac{Re^{i\theta} - \alpha}{Re^{-i\theta} - \bar{\alpha}} \right| = \frac{|\mu|}{R} = \frac{R^2}{R} = R.$$

Therefore (10.35) maps C *into* C, and hence *onto* C, by the same argument as before. This example is considered further in Problems 10.18–10.22.

*50. Lobachevskian Geometry

In the preceding section, we saw that every function of the form

$$w = L(z) = \mu \frac{z - \alpha}{R^2 - \bar{\alpha}z} \qquad (|\alpha| < R, \quad |\mu| = R^2) \tag{10.36}$$

transforms the disk K: $|z| < R$ into itself. In particular, the function

$$\bar{\mu} \frac{z - \bar{\alpha}}{R^2 - \alpha z},$$

which we denote by $\bar{L}(z)$, has the same property, and obviously $\overline{L(z)} = \bar{L}(\bar{z})$. Moreover, it is easy to see that by letting α and μ in (10.36) take all possible

values satisfying the conditions $|\alpha| < R$ and $|\mu| = R^2$, we obtain all possible Möbius transformations carrying the disk K into itself. In fact, any such transformation $w = L(z)$ must carry some point $\alpha \in K$ into the center of K, and hence, according to Sec. 49, $L(z)$ must have the form (10.36).

If L_1 and L_2 are two Möbius transformations carrying K into itself, their product L_1L_2 (or L_2L_1) is also a Möbius transformation carrying K into itself, and hence is of the form (10.36). Moreover, the unit transformation and the inverse of any transformation (10.36) are obviously Möbius transformations carrying K into itself, and hence are also of the form (10.36). Therefore the set of all transformations of the form (10.36), with fixed $R > 0$, forms a group, which we denote by $\mathscr{K}$. The group $\mathscr{K}$ is a *subgroup* of the group $\mathscr{M}$ of all Möbius transformations.[13] Moreover, as we now show, $\mathscr{K}$ *has a remarkable geometric interpretation, i.e.,* $\mathscr{K}$ *can be regarded as the group of motions of the plane in Lobachevskian geometry.*

First, we introduce some terminology. The disk K itself will be called the *Lobachevskian plane* (briefly, the $\mathscr{L}$*-plane*), the points of K will be called *Lobachevskian points* (briefly, $\mathscr{L}$*-points* or just *points*), and circular arcs or line segments contained in K and *orthogonal* to the circle $\Gamma: |z| = R$ will be called *Lobachevskian lines* (briefly, $\mathscr{L}$*-lines*). As we now show, the $\mathscr{L}$-plane and its points and lines satisfy many of the same relations as the usual Euclidean plane and its points and lines (thereby justifying our nomenclature). In fact, the following propositions are valid in the $\mathscr{L}$-plane:[14]

Axioms of Incidence

I_1. *For every pair of points A and B, there exists an* $\mathscr{L}$*-line passing through A and B.*

I_2. *Given two points A and B, there exists no more than one* $\mathscr{L}$*-line passing through A and B.*

I_3. *There exist at least two points on a given* $\mathscr{L}$*-line. Moreover, there exist at least three points which do not lie on a given* $\mathscr{L}$*-line.*

Axioms of Order

II_1. *If A, B and C are points of the same* $\mathscr{L}$*-line, and if B lies between A and C, then B lies between C and A.*

II_2. *If A and C are points of a given* $\mathscr{L}$*-line, then the line contains at least one point B such that C lies between A and B.*

---

[13] For a discussion of the concept of a *subgroup*, see e.g., G. Birkhoff and S. MacLane, *op. cit.*, Chap. 6, Sec. 8, and V. I. Smirnov, *op. cit.*, Sec. 63.

[14] The five groups of axioms to be given in this section are known collectively as *Hilbert's axioms*. See D. Hilbert, *Foundations of Geometry* (translated by E. J. Townsend), The Open Court Publishing Co., La Salle, Illinois (1947).

II_3. *Among any three points of an $\mathscr{L}$-line, no more than one point lies between two others.*

II_4. *Let A, B and C be three points which do not lie on the same $\mathscr{L}$-line, and let l be an $\mathscr{L}$-line which does not pass through any of the points A, B and C. Then if l passes through a point of the segment AB, l must pass through either a point of the segment AC or a point of the segment BC.*[15]

We now prove Axioms I_1 and I_2, leaving the proof of the remaining axioms as exercises for the reader.[16] We begin by assuming that one of the two given points, say A, coincides with the center of the disk K. Then the diameter of K passing through B is the $\mathscr{L}$-line satisfying Axiom I_1. Moreover, suppose that contrary to Axiom I_2, there exists another $\mathscr{L}$-line passing through the points A and B. Then this line must be a circular arc l orthogonal to Γ, and hence the radii of Γ drawn to the end points of l must be tangent to l. But this is impossible, since each of the radii shares two different points with l (the point A and one end point of l).

Next suppose that neither of the points A and B coincides with the center of K. In this case, making a transformation of the form (10.36), where α is the complex number corresponding to A, we can reduce the problem to the case just considered, and then, making the inverse transformation, we can find the unique $\mathscr{L}$-line passing through the given points (the image under the inverse transformation of the diameter of K passing through A and B).

By a *Lobachevskian motion* (briefly, an *$\mathscr{L}$-motion*), we mean any Möbius transformation of the form (10.36), i.e., a Möbius transformation carrying the disk K into itself. Every $\mathscr{L}$-motion is a one-to-one mapping of the $\mathscr{L}$-plane onto itself, such that points go into points and $\mathscr{L}$-lines into $\mathscr{L}$-lines (here we use the circle-preserving property and the conformality of Möbius transformations).

Now consider a set E of points in the $\mathscr{L}$-plane, and let $\bar{E}$ be the set of points symmetric to the points of E with respect to the real axis, i.e., the set of complex conjugates of the points of E. Moreover, let F be any other set of points in the $\mathscr{L}$-plane. Then we introduce the following definition: *The set E is congruent to the set F in the Lobachevskian sense (briefly, E is $\mathscr{L}$-congruent to F) if and only if there exists an $\mathscr{L}$-motion L transforming F into E or into $\bar{E}$, i.e., such that $L(F) = E$ or $L(F) = \bar{E}$.*

---

[15] The following conventions are in force in this section: Segments of $\mathscr{L}$-lines (i.e., circular arcs or straight line segments) will be written simply as AB, $\alpha\beta$, etc., without using the symbol $\frown$ or an overbar. Moreover, by the set $\bar{E}$ we shall mean the set of complex conjugates of the points in E, and not the closure of E.

[16] The reader should not be disturbed by the fact that we are proving axioms! From our point of view, all these axioms are *theorems* of Euclidean geometry, involving a disk $K: |z| < R$, points of K, and circular arcs or line segments contained in K. In proving these axioms, we have a right to use any and all concepts and propositions from Euclidean geometry.

Next we show that the relation of $\mathscr{L}$-congruence, denoted by the symbol $\equiv$, is an *equivalence relation*,[17] which means that it has the following properties:

1. *Symmetry*: $E \equiv E$;
2. *Reflexivity*: $E \equiv F$ implies $F \equiv E$;
3. *Transitivity*: $E \equiv F$ and $F \equiv G$ imply $E \equiv G$.

The proof of property 1 is obvious (note also that $E \equiv \bar{E}$). To prove property 2, we observe that if $L(F) = E$, then $L^{-1}(E) = F$, while if $L(F) = \bar{E}$, then $\bar{L}^{-1}(E) = \bar{F}$ [recall that $\overline{L(E)} = \bar{L}(\bar{E})$]. Finally, to prove property 3, we note that $L_1(F) = E$ or $L_1(F) = \bar{E}$, which together with $L_2(G) = F$ or $L_2(G) = \bar{F}$, implies one of the formulas

$$L_1L_2(G) = E, \quad L_1L_2(G) = \bar{E}, \quad \bar{L}_1L_2(G) = E, \quad \bar{L}_1L_2(G) = \bar{E},$$

each of which amounts to the assertion that $E \equiv G$.

We now state and prove another set of axioms:

Axioms of Congruence

III_1. *If A and B are two points of an $\mathscr{L}$-line l, and if A' is a point of another $\mathscr{L}$-line l' (which may coincide with l), then one and only one point B' can always be found on l', on a given side of A', such that the segment $A'B'$ is $\mathscr{L}$-congruent to the segment AB.*

III_2. *If $AB \equiv A'B'$ and $AB \equiv A''B''$, then $A'B' \equiv A''B''$.*

III_3. *Let AB and BC be two segments of an $\mathscr{L}$-line l, without common interior points, and let $A'B'$ and $B'C'$ be two segments of another $\mathscr{L}$-line l' (which may coincide with l), also without common interior points. Then $AC \equiv A'C'$ if $AB \equiv A'B'$ and $BC \equiv B'C'$.*

Before stating two additional axioms of congruence we prove those just enumerated.

To prove Axiom III_1, we let C denote one end point of the circular arc (or line segment) l representing the given $\mathscr{L}$-line. According to Problem 10.18, we can map the disk K onto itself in such a way that the point A goes into O, the center of K, while the point C goes into the point R of the circumference of K (see Figure 10.3). As a result, we obtain an $\mathscr{L}$-motion L which transforms l into the $\mathscr{L}$-line represented by the diameter d of the disk K which lies on the real axis, with the $\mathscr{L}$-ray (i.e., $\mathscr{L}$-half-line) AC going into the $\mathscr{L}$-ray OR. If AC contains the segment AB, then the image $B_1 = L(B)$ will lie on OB. On the $\mathscr{L}$-line l' we locate the $\mathscr{L}$-ray $A'C'$ lying on the given side of A'. Let L' be the $\mathscr{L}$-motion transforming l' into d and $A'C'$ into OR, and let B' be the point of the $\mathscr{L}$-line l' such that $L'(B') = B_1$. Then it is obvious that the segment $A'B'$ of the $\mathscr{L}$-line l' lies on the appropriate side

[17] See e.g., G. Birkhoff and S. MacLane, *op. cit.*, p. 31.

of A' and that the $\mathscr{L}$-motion $L^{-1}L'$ transforms l' into l and $A'B'$ into AB. Therefore $A'B' \equiv AB$, as asserted.

We still have to show that B' is the *unique* point satisfying the conditions of the axiom. Suppose there is another point $B'' \neq B'$ on the $\mathscr{L}$-ray $A'C'$ such that the segment $A'B''$ is $\mathscr{L}$-congruent to AB (again see Figure 10.3). Then on the $\mathscr{L}$-ray OR we can find a point $B_2 \neq B_1$, where $B_2 = L'(B'')$, such that $OB_2 \equiv OB_1$. By definition, this means that there exists an $\mathscr{L}$-motion Λ such that $OB_2 = \Lambda(OB_1)$.[18] In other words, $w = \Lambda(z)$ is a transformation of the disk K into itself, which carries the line segment OB_1 into OB_2, and hence carries the diameter d of K lying along the real axis into itself. *A priori*, there are two possibilities:

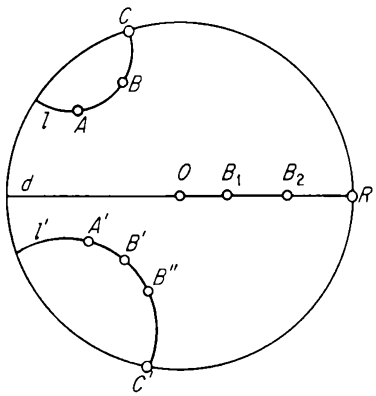

FIGURE 10.3

1. The point O goes into O, and the point B_1 goes into B_2. Then, using formula (10.36) with $\alpha = 0$, we see that $\Lambda(z)$ must have the form

$$w = \Lambda(z) = \frac{\mu}{R^2} z.$$

But this implies $|w| = |z|$, since $|\mu| = R^2$, which contradicts the assumption that B_1 goes into B_2 (since $|B_1| \neq |B_2|$).

2. The point B_1 goes into O, and the point O goes into B_2. In this case, setting $\alpha = B_1$ in (10.36), we have

$$w = \Lambda(z) = \mu \frac{z - B_1}{R^2 - B_1 z}$$

and

$$B_2 = \Lambda(0) = -\frac{\mu B_1}{R^2}.$$

But this implies

$$|B_2| = \frac{\mu}{R^2} |B_1| = |B_1|,$$

which again contradicts the condition $|B_2| \neq |B_1|$, thereby completing the verification of Axiom III_1.

---

[18] There is no need to consider the set $\overline{OB_2}$ symmetric to OB_2 with respect to the real axis, since $\overline{OB_2}$ coincides with OB_2.

Axiom III_2 is an immediate consequence of the symmetry and transitivity of $\mathscr{L}$-congruence.

To verify Axiom III_3, we argue as follows: We make an $\mathscr{L}$-motion L which transforms l into the diameter d in such a way that the point B goes into O, the center of K, and the point C goes into a point of the positive real axis. Then, according to the conditions of the axiom, A lies on the negative real axis. Suppose L transforms A and C into A_1 and C_1 (see Figure 10.4). If we consider the analogous $\mathscr{L}$-motion L' constructed by starting from the $\mathscr{L}$-line l', we obtain a point A_2 on the negative real axis, the point $B = O$, and a point C_2 on the positive real axis. It follows from $AB \equiv A'B'$, $AB \equiv A_1B_1$ and $A'B' \equiv A_2B_2$ that $A_1B_1 \equiv A_2B_2$, i.e., $A_1O \equiv A_2O$, and similarly that $OC_1 \equiv OC_2$. But A_1 and A_2 lie on the same side of O, and hence $A_1O \equiv A_2O$ implies $A_1 = A_2$. In the same way, $OC_1 \equiv OC_2$ implies $C_1 = C_2$. Therefore $A_1C_1 \equiv A_2C_2$, and this, together with the transitivity of $\mathscr{L}$-congruence, implies $AC \equiv A'C'$, as required.

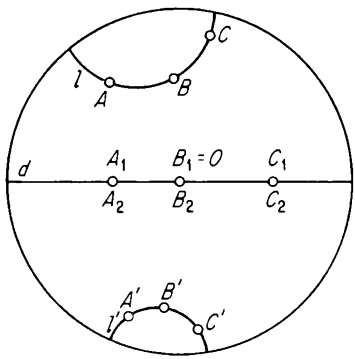

FIGURE 10.4

Next we introduce the concept of an $\mathscr{L}$-angle (as a geometric figure). By an *$\mathscr{L}$-angle* we mean a pair of $\mathscr{L}$-rays (i.e., $\mathscr{L}$-half-lines) h and k emanating from a common point A and contained in different $\mathscr{L}$-lines. The angle in question is denoted by $\angle(h, k)$ [or $\angle(k, h)$, depending on our convention as to direction], the point A is called the *vertex* of $\angle(h, k)$, and the $\mathscr{L}$-rays h and k are called the *sides* of $\angle(h, k)$ [see Figure 10.5]. Obviously, the sides of any $\mathscr{L}$-angle divide the $\mathscr{L}$-plane into two domains such that any pair of points belonging to one of the domains (the domain[19] indicated by the shaded area in Figure 10.5) can be joined by a segment of an $\mathscr{L}$-line (i.e., by a circular arc or line segment orthogonal to the circle Γ: $|z| = R$) which does not intersect the sides h and k, whereas there exist pairs of points belonging to the other domain (the domain that is not $\mathscr{L}$-convex) which cannot be joined by any such segment. (The best way of seeing this is to make an $\mathscr{L}$-motion carrying the vertex of A into the origin and the sides h and k into radii of the circle Γ.) The first domain is called the *interior* of $\angle(h, k)$, and the other domain is called the *exterior* of $\angle(h, k)$.

We now state and prove two additional congruence axioms, called *axioms of planar congruence*:[20]

---

[19] This domain might be called *$\mathscr{L}$-convex*.

[20] The axioms of congruence considered previously are called *axioms of linear congruence*.

III$_4$. *Let* $\angle(h, k)$ *be an* $\mathscr{L}$*-angle, let* l' *be an* $\mathscr{L}$*-line, and let* $h' \subset l'$ *be an* $\mathscr{L}$*-ray emanating from a point* $A \in l'$. *Then, on a given side of* l', *there exists one and only one ray* k' *emanating from* A *such that the angle* $\angle(h', k')$ *is congruent to the angle* $\angle(h, k)$.[21] *In particular, every angle is congruent to itself, i.e.,* $\angle(h, k) \equiv \angle(h, k)$ [*and also* $\angle(h, k) \equiv \angle(k, h)$].

The statement and proof of Axiom III$_4$ resembles that of Axiom III$_1$. However to prove the uniqueness of the ray k', we now make use of the fact that the size of an $\mathscr{L}$-angle (its "measure") is preserved under $\mathscr{L}$-motions (this follows from the conformality of Möbius transformations). Then it is obvious that two angles $\angle(h', k')$ and $\angle(h', k'')$ with a common vertex A and a common side h', such that k' and k'' lie on the same side of l' but represent different rays, cannot be $\mathscr{L}$-congruent.

Before stating our last axiom of congruence, we introduce some more definitions: Three segments AB, BC and CA of three different $\mathscr{L}$-lines are said to form an *$\mathscr{L}$-triangle*, with *vertices* A, B, C and *sides* AB, BC, CA. If h and k are rays emanating from A and containing the sides AB and AC, respectively, the angle $\angle(h, k)$ is called the angle of the triangle ABC *between* the sides AB and AC, and *opposite* the side BC. This angle, whose interior contains all the interior points of the triangle ABC, is denoted by $\angle BAC$ or simply $\hat{A}$.

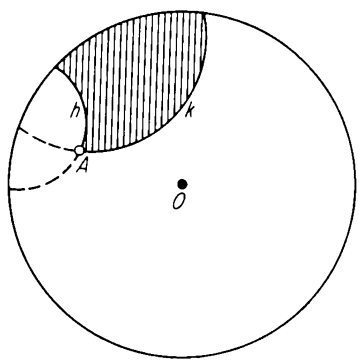

FIGURE 10.5

III$_5$. *Given two triangles* ABC *and* $A'B'C'$, *if* $AB \equiv A'B'$, $AC \equiv A'C$ *and* $\angle BAC \equiv \angle B'A'C'$, *then* $\angle ABC \equiv \angle A'B'C'$ *and* $\angle ACB \equiv \angle A'C'B'$.

To verify this last axiom of congruence, we subject the triangles ABC and $A'B'C'$ to $\mathscr{L}$-motions carrying the vertices A and A' into the origin O and the sides AB and $A'B'$ into the same segment OB_1 of the diameter d (the configuration resembles that of Figure 10.3). Then the sides AB, $A'B'$, AC and $A'C'$ are represented by segments of radii of the circle $\Gamma: |z| = R$ with common vertex O, where $O = A = A'$. If AC and $A'C'$ lie on the same side of the diameter d of Γ containing the segment $AB = A'B'$, then, since the angles $\angle BAC$ and $\angle B'A'C'$ are congruent, AC and $A'C'$ must lie on the same $\mathscr{L}$-ray emanating from O. Therefore the points C and C' coincide

[21] From now on, to avoid monotony, we shall often drop the prefix $\mathscr{L}$ in expressions like $\mathscr{L}$-ray, $\mathscr{L}$-angle, $\mathscr{L}$-congruent, $\mathscr{L}$-triangle, etc.

(since $AC \equiv A'C'$). It follows that the segments AC and $A'C'$ coincide, and so do the segments BC and $B'C'$, so that $\angle ABC \equiv \angle A'B'C'$ and $\angle ACB \equiv \angle A'C'B'$, as required. (This also means that the triangles ABC and $A'B'C'$ themselves are congruent.) Finally, we note that if AC and $A'C'$ lie on opposite sides of the diameter d, then a preliminary reflection of one of these segments in d reduces the problem to the case just considered.

We now introduce the concept of the *size of an angle* and the *length of a segment*. These concepts, which we designate by the same term *measure*, must satisfy the following conditions:

1. *Measure must be a nonnegative real number.*

2. *Measure must be invariant under $\mathscr{L}$-motions, i.e., congruent angles and segments must have the same measure.*

3. *Measure must be additive in the sense that* i) *the measure of a segment AC must equal the sum of the measures of the segments AB and BC if B is any interior point of AC,*[22] *and* ii) *the measure of an angle $\angle(h, k)$ must equal the sum of the measures of the angles $\angle(h, l)$ and $\angle(l, k)$ if l is any ray emanating from the vertex of $\angle(h, k)$ and contained in the interior of $\angle(h, k)$.*

As far as measure of angles is concerned, the situation is particularly simple. We need only choose the measure of the $\mathscr{L}$-angle $\angle(h, k)$ to be the ordinary *Euclidean* measure of the angle between the circular arcs (or line segments) representing h and k. Then conditions 1 and 3 are obviously satisfied, and condition 2 is also satisfied because of the conformality of $\mathscr{L}$-motions (i.e., of Möbius transformations).

However, the situation concerning measure of segments is more complicated. Although the Euclidean length of the circular arc (or line segment) representing a segment AB of an $\mathscr{L}$-line also satisfies conditions 1 and 3, it does not satisfy condition 2, since $\mathscr{L}$-motions (except for rotations about the point O and reflections in a diameter) change Euclidean length. To find a measure satisfying all these conditions, we use the invariance under $\mathscr{L}$-*motions* of the *cross ratio* of four points lying on an arbitrary Euclidean line or circle (see Sec. 46). Thus let a and b be the complex numbers corresponding to the end points of a segment AB of some $\mathscr{L}$-line l, and let α and β be the complex numbers corresponding to the end points of the circular arc (or line segment) containing AB. Suppose that α and β are labeled in such a way that the four points, α, β, a, b appear on the circular arc in the order α, a, b, β. Under an $\mathscr{L}$-motion, these points are transformed into new points α', a', b', β', where the points α' and β' stay on the circle $\Gamma: |z| = R$, and the relative order of

[22] I.e., any point of AC other than the end points A and C.

the points α', a', b', β' on the new arc (into which the original arc is transformed by the $\mathscr{L}$-motion) is not destroyed. Moreover, the cross ratio

$$(\alpha, \beta, b, a) = \frac{b-\alpha}{b-\beta} : \frac{a-\alpha}{a-\beta} \tag{10.37}$$

remains unchanged.

We now show that (10.37) is a positive number greater than 1. First we perform the $\mathscr{L}$-motion which transforms the line l containing AB into the diameter d lying along the real axis, in such a way that the point a goes into the origin O and the point α goes into the point R (cf. Problem 10.18). Then β goes into $-R$ and b goes into a point b' such that $-R < b' < 0$ (this last fact follows from the preservation of relative order of points on an $\mathscr{L}$-line under an $\mathscr{L}$-motion). Therefore we find that

$$(\alpha, \beta, b, a) = (R, -R, b', 0) = \frac{b'-R}{b'+R} : \frac{0-R}{0+R} = \frac{R-b'}{R+b'} > 1,$$

as asserted, and hence, as a measure of the segment AB, the cross ratio (10.37) satisfies conditions 1 and 2. However, it is easy to see that (10.37) does not satisfy the additivity condition 3. In fact, let C be an interior point of AB, and let the corresponding complex number be c. Then

$$(\alpha, \beta, c, a) = \frac{c-\alpha}{c-\beta} : \frac{a-\alpha}{a-\beta}, \qquad (\alpha, \beta, b, c) = \frac{b-\alpha}{b-\beta} : \frac{c-\alpha}{c-\beta},$$

so that

$$\begin{aligned}(\alpha, \beta, b, a) = \frac{b-\alpha}{b-\beta} : \frac{a-\alpha}{a-\beta} &= \left(\frac{c-\alpha}{c-\beta} : \frac{a-\alpha}{a-\beta}\right)\left(\frac{b-\alpha}{b-\beta} : \frac{c-\alpha}{c-\beta}\right) \\ &= (\alpha, \beta, c, a)(\alpha, \beta, b, c).\end{aligned}$$

Thus the cross ratio (α, β, b, a) does not equal the sum of the cross ratios (α, β, c, a) and (α, β, b, c), but rather their product (which in general does not equal the sum).

It is now clear that we can satisfy all the requirements imposed on a measure by choosing the *$\mathscr{L}$-length* of the segment AB to be the *logarithm* of the cross ratio, i.e., the quantity

$$\ln (\alpha, \beta, b, a).$$

In fact, since $(\alpha, \beta, b, a) > 1$, we have $\ln (\alpha, \beta, b, a) > 0$, and moreover, since (α, β, b, a) is invariant under $\mathscr{L}$-motions, so is $\ln (\alpha, \beta, b, a)$. Finally, as just shown, we have

$$\ln (\alpha, \beta, b, a) = \ln (\alpha, \beta, c, a) + \ln (\alpha, \beta, b, c)$$

for every point C of the segment AB. It should be noted that $\ln (\alpha, \beta, b, a)$ vanishes when $a = b$, i.e., when $A = B$, as is to be expected.

Remark. With this definition of the $\mathscr{L}$-length of the segment AB (or equivalently, of the *$\mathscr{L}$-distance* between the points A and B), the points of

the circle Γ, which is the boundary of the disk K, can be regarded as *points at infinity in the $\mathscr{L}$-plane.* To see this, we fix the point A on the line l and make the point B approach Γ along l in the direction away from A.[23] Then $b \to \beta$ and both the cross ratio (10.37) and its logarithm approach $+\infty$.

Continuing our verification of Hilbert's system of axioms, we temporarily omit Group IV (the *axiom of parallels*) and pass directly to Group V (the *axioms of continuity*):

V_1 (*Archimedes' axiom*). *Let A_1 be an arbitrary point of a given $\mathscr{L}$-line l such that A_1 lies between two given points A and B of l. Construct a sequence of points $A_2, A_3, A_4, \ldots$ such that A_1 lies between A and A_2, A_2 lies between A_1 and A_3, A_3 lies between A_2 and A_4, etc., and such that all the segments AA_1, A_1A_2, A_2A_3, $A_3A_4, \ldots$ are congruent. Then in the sequence of points $A_2, A_3, A_4, \ldots$ there always exists a point A_n such that the point B lies between A and A_n.*

To verify this axiom, we need only note that according to properties 2 and 3 of $\mathscr{L}$-measure, the length of the segment AA_n is n times the length of the segment AA_1. Therefore the length of the segment AA_n grows arbitrarily large as n increases, and hence as $n \to \infty$, A_n approaches one end point of the circular arc (or line segment) representing the given $\mathscr{L}$-line. Since the point B lies between A and this end point, then, starting from some value of n, B lies between A and A_n.

As our second axiom of continuity, we choose *Cantor's axiom of continuity* (instead of *Hilbert's axiom of completeness*[24]):

V_2. *Given two sequences of points $A_1, A_2, \ldots, A_n, \ldots$ and $B_1, B_2, \ldots, B_n, \ldots$ on a given $\mathscr{L}$-line l, suppose that B_q lies between A_p and B_{q-1} while A_p lies between A_{p-1} and B_q, for arbitrary p and q. Then there is at least one point $C \in l$ lying between A_p and B_q for arbitrary p and q.*

We can immediately convince ourselves of the validity of Axiom V_2 by regarding it as a theorem of Euclidean geometry, which expresses a property of the circular arc or line segment representing the $\mathscr{L}$-line l.

So far, we have observed a complete resemblance between Euclidean geometry and Lobachevskian geometry. This resemblance is expressed in the fact that all the axioms of incidence, order, congruence and continuity have the same formulation in both geometries. It follows that all the theorems of Euclidean geometry with proofs based on these axioms remain valid in Lobachevskian geometry. The difference between the two geometries only appears when we consider Axiom IV (the *axiom of parallels*):

---

[23] Alternatively, we can fix B and make A approach Γ along l in the direction away from B, so that $a \to \alpha$.

[24] See D. Hilbert, *op. cit.*, p. 25.

IV. *Given a point A of the $\mathscr{L}$-plane and an $\mathscr{L}$-line l such that A does not lie on l, an infinite set of different $\mathscr{L}$-lines which have no points in common with l can be drawn through A.*[25]

A glance at Figure 10.6 is enough to convince us of the truth of Axiom IV. Through the point A we have drawn two arcs $A\alpha$ and $A\beta$ orthogonal to the circle Γ, each of which has an end point in common with the arc $\alpha\beta$ representing a certain $\mathscr{L}$-line l. The arcs $A\alpha$ and $A\beta$ represent $\mathscr{L}$-lines which have no finite points in common with l (recall that only points *inside* the circle Γ are finite points of the $\mathscr{L}$-plane). These two $\mathscr{L}$-lines are called the *$\mathscr{L}$-parallels* to l drawn through the point A, and are characterized by the fact that they share a point at infinity with l (α in one case, β in the other). Obviously, every $\mathscr{L}$-line passing through A and lying in the shaded area shown in Figure 10.6 (e.g., the arc $\gamma\delta$) has no points at all in common with l, not even points at infinity. Therefore $\mathscr{L}$-lines like $\gamma\delta$ are not called $\mathscr{L}$-parallels.

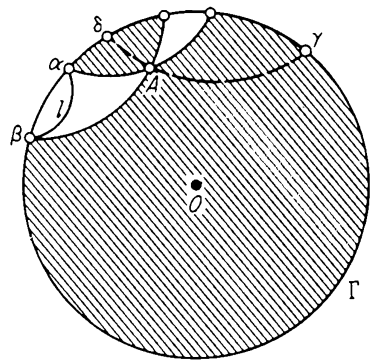

FIGURE 10.6

We now prove two theorems of Lobachevskian geometry which differ from their Euclidean counterparts:

THEOREM 10.9. *The sum of the angles of an $\mathscr{L}$-triangle is less than π.*

Proof. We subject an arbitrary $\mathscr{L}$-triangle ABC, with angles α, β and γ, to the $\mathscr{L}$-motion carrying the vertex A into the center of the disk K. Then ABC takes the form shown in Figure 10.7. Obviously, the arc $B'C'$ which represents the $\mathscr{L}$-line containing the side BC of the $\mathscr{L}$-triangle is convex toward the vertex A. In fact, A is the point of intersection of the tangents to the arc $B'C'$ drawn at the points B' and C', and hence A lies outside the circle containing the arc $B'C'$. Joining B and C to A by (Euclidean) line segments, we obtain the Euclidean triangle ABC, the sum of whose angles is

$$\hat{A} + \hat{B} + \hat{C} = \pi.$$

But $\alpha = \hat{A}$, $\beta < \hat{B}$, and $\gamma < \hat{C}$, and hence

$$\alpha + \beta + \gamma < \pi,$$

as asserted.

THEOREM 10.10. *Two $\mathscr{L}$-triangles with equal angles are $\mathscr{L}$-congruent.*

---

[25] The corresponding axiom of Euclidean geometry goes as follows: *Given a point A of the Euclidean plane and a line l such that A does not lie on l, one and only one line which has no points in common with l can be drawn through A.*

Proof. Let ABC and $A_1B_1C_1$ be two $\mathscr{L}$-triangles with equal angles, i.e., let $\alpha = \alpha_1$, $\beta = \beta_1$ and $\gamma = \gamma_1$. We subject these two triangles to the $\mathscr{L}$-motions carrying the points A and A_1 into the center of the disk K and carrying the sides AB and A_1B_1 into segments contained in the same radius h of the circle Γ (of course, the two $\mathscr{L}$-motions will be different). If as a result, the sides AC and A_1C_1 do not lie on the same radius of the circle, then, since the angles α and α_1 are equal, the sides AC and A_1C_1 lie on radii which are symmetric with respect to h. Therefore, by reflecting one of the triangles in h, we can make the radii containing the sides AC and A_1C_1 coincide. Then, if the triangles themselves do not coincide, we encounter one of the two cases shown in Figure 10.8.

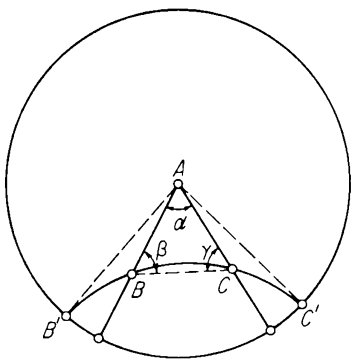

FIGURE 10.7

In the case shown in Figure 10.8(a), the sides BC and B_1C_1 intersect in a point D, thereby forming two $\mathscr{L}$-triangles BDB_1 and CDC_1 (one of which degenerates into a point if D coincides with B and B_1, or with C and C_1). Calculating the angles of the $\mathscr{L}$-triangle BDB_1, we find that

$$\hat{B}_1 = \beta_1, \quad \hat{B} = \pi - \beta, \qquad \hat{D} > 0,$$

and hence

$$\hat{B}_1 + \hat{B} + \hat{D} = \pi + \beta_1 - \beta + D = \pi + D,$$

since $\beta_1 = \beta$. Thus we arrive at the conclusion that the sum of the angles of the $\mathscr{L}$-triangle BDB_1 is greater than π, which contradicts Theorem 10.9.

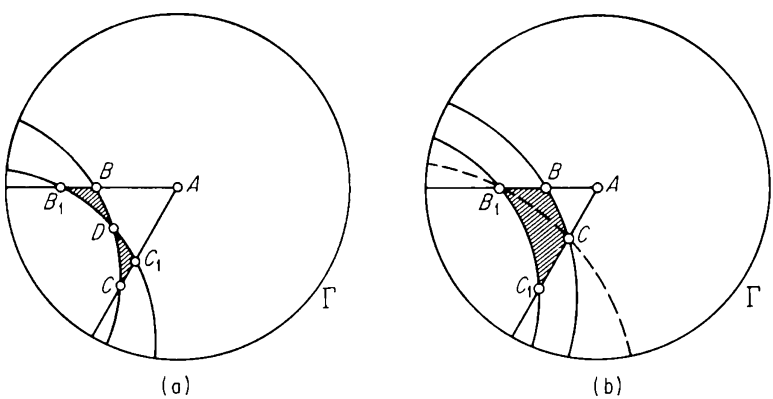

FIGURE 10.8

In the case shown in Figure 10.8(b), the sides BC and B_1C_1 do not intersect. Then the angles of the $\mathscr{L}$-quadrilateral BB_1C_1C are

$$\hat{B} = \pi - \beta, \quad \hat{B}_1 = \beta_1, \quad \hat{C} = \pi - \gamma, \quad \hat{C}_1 = \gamma_1,$$

and hence

$$\hat{B} + \hat{B}_1 + \hat{C} + \hat{C}_1 = 2\pi + \beta_1 - \beta + \gamma_1 - \gamma = 2\pi.$$

But the sum of the angles of this $\mathscr{L}$-quadrilateral can be calculated by drawing the $\mathscr{L}$-diagonal B_1C dividing BB_1C_1C into two $\mathscr{L}$-triangles BCB_1 and B_1CC_1, and then finding the sum of the angles of these two triangles. Since, according to Theorem 10.9, this sum must be less than 2π, we again arrive at a contradiction. Therefore the indicated $\mathscr{L}$-motions must cause the $\mathscr{L}$-triangles ABC and $A_1B_1C_1$ to coincide, i.e., these triangles are $\mathscr{L}$-congruent, as asserted.

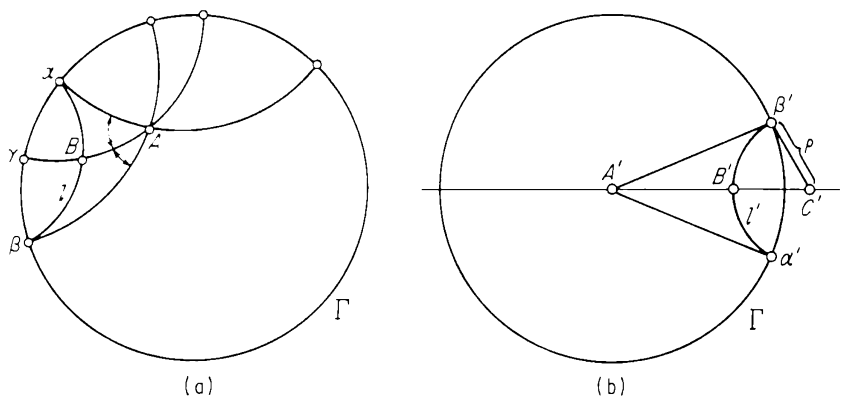

FIGURE 10.9

Let l be an $\mathscr{L}$-line, and let A be a point which does not belong to l. Through A we draw the two $\mathscr{L}$-parallels to l, and we drop the $\mathscr{L}$-perpendicular AB from A onto l, as shown in Figure 10.9(a). Suppose we make the $\mathscr{L}$-motion carrying A into A', the origin of the disk, and carrying AB into a segment $A'B'$ of the positive real axis, as shown in Figure 10.9(b). Then the $\mathscr{L}$-parallels to l go into radii of the circle Γ, and l itself goes into the circular arc l'. From the symmetry of the figure we see that the two angles $\angle \alpha AB \equiv \angle \alpha' A'B'$ and $\angle BA\beta \equiv \angle B'A'\beta'$ are equal. Their common value, called the *angle of parallelism*, is of great importance in Lobachevskian geometry.

THEOREM 10.11. *Let δ be the $\mathscr{L}$-distance from the point A to the line l, i.e., the $\mathscr{L}$-length of the perpendicular AB dropped from A onto l. Then the angle of parallelism is a function of δ only, given by*

$$\Pi(\delta) = \arctan \frac{1}{\sinh \delta}. \tag{10.38}$$

Proof. We have

$$\delta = \ln(-R, R, b', 0) = \ln\left(\frac{b' + R}{b' - R} : \frac{0 + R}{0 - R}\right) = \ln\frac{R + b'}{R - b'},$$

where $b' > 0$ is the abscissa of the point B'. It follows that

$$\frac{R + b'}{R - b'} = e^{\delta}, \qquad b' = R\frac{e^{\delta} - 1}{e^{\delta} + 1}.$$

Suppose the tangent to Γ at the point β' intersects the real axis in the point C' [see Figure 10.9(b)]. Then C' is obviously the center of the circle containing the arc l', and $C'\beta' = \rho$ is the radius of this circle.[26] Moreover, $A'\beta'$ is tangent to the arc $\alpha'\beta'$ at β', and hence

$$(A'\beta')^2 = A'B'(A'B' + 2\rho). \tag{10.39}$$

Solving (10.39) for ρ, we obtain

$$\rho = \frac{R^2 - b'^2}{2b'} = \frac{R^2 - R^2[(e^{\delta} - 1)/(e^{\delta} + 1)]^2}{2R(e^{\delta} - 1)/(e^{\delta} + 1)} = \frac{2Re^{\delta}}{e^{2\delta} - 1}$$
$$= \frac{R}{(e^{\delta} - e^{-\delta})/2} = \frac{R}{\sinh\delta}.$$

Finally, examining the triangle $A'\beta'C'$, we find that

$$\Pi(\delta) = \arctan\frac{C'\beta'}{A'\beta'} = \arctan\frac{\rho}{R} = \arctan\frac{1}{\sinh\delta},$$

as asserted.

Remark. It follows from (10.38) that the angle of parallelism satisfies the inequality

$$0 < \Pi(\delta) < \frac{\pi}{2}.$$

Moreover, as $\delta \to 0$,

$$\sinh\delta \to 0, \quad \frac{1}{\sinh\delta} \to +\infty, \quad \Pi(\delta) \to \frac{\pi}{2},$$

while as $\delta \to \infty$,

$$\sinh\delta \to +\infty, \quad \frac{1}{\sinh\delta} \to 0, \quad \Pi(\delta) \to 0.$$

Formula (10.38) plays a basic role in Lobachevskian *trigonometry*.

---

[26] Here, we also use $C'\beta'$, $A'\beta'$ and $A'B'$ to denote the lengths of the corresponding *Euclidean* line segments.

For the further development of the material presented in this section, we refer the reader to the extensive literature on non-Euclidean geometry.[27]

51. The Mapping $w = \frac{1}{2}\left(z + \frac{1}{z}\right)$

A mapping by a rational function of order higher than 1 was studied in Sec. 37, in connection with the function

$$w = (z - a)^n \qquad (n > 1).$$

However, this function is not meromorphic in the full sense of the word, since it is actually entire. We now study the rational function

$$w = \lambda(z) = \frac{1}{2}\left(z + \frac{1}{z}\right) = \frac{z^2 + 1}{2z}, \tag{10.40}$$

which comes up in the course of solving a variety of problems. In fact, because of the use which the Russian scientist Joukowski (1847–1921) made of this function in aerodynamics, it is often referred to as the *Joukowski function*. Obviously, $w = \lambda(z)$ is a rational function of order 2, which does not reduce to an entire function and which satisfies the condition

$$\lambda(z) = \lambda\left(\frac{1}{z}\right). \tag{10.41}$$

It follows from (10.41) that under the mapping $w = \lambda(z)$ every point of the w-plane except $w = \pm 1$ has two (and only two) distinct inverse images z_1 and z_2 satisfying the relation

$$z_1 z_2 = 1. \tag{10.42}$$

Now let γ be the unit circle $|z| = 1$, with interior $I(\gamma)$ and exterior $E(\gamma)$. Then, according to (10.42), $z_1 \in I(\gamma)$ if and only if $z_2 \in E(\gamma)$. Moreover, the two sets $\lambda[I(\gamma)]$ and $\lambda[E(\gamma)]$ are identical. Clearly, the function $w = \lambda(z)$ is continuous (in the wide sense) on the closed domains $\overline{I(\gamma)}$ and $\overline{E(\gamma)}$, and one-to-one on $I(\gamma)$ and $E(\gamma)$. Therefore, according to Theorem 6.1, $w = \lambda(z)$ maps $I(\gamma)$ and $E(\gamma)$ onto some domain $\mathcal{G}$ in the w-plane. Moreover, according to Theorem 6.3, to determine the boundary of this domain, we have to find the image $\Gamma = \lambda(\gamma)$ of the unit circle. But if

$$z = e^{it} \qquad (0 \leqslant t \leqslant 2\pi),$$

then

$$w = u + iv = \tfrac{1}{2}(e^{it} + e^{-it}) = \cos t \qquad (0 \leqslant t \leqslant 2\pi),$$

---

[27] See e.g., K. Borsuk and W. Smielew, *Foundations of Geometry*, revised edition (translated by E. Marquit), North-Holland Publishing Co., Amsterdam (1960); N. V. Efimov, *Höhere Geometrie*, VEB Deutscher Verlag der Wissenschaften, Berlin (1960); H. E. Wolfe, *Introduction to Non-Euclidean Geometry*, Holt, Rinehart and Winston, Inc., New York (1945).

i.e., Γ is the segment $-1 \leqslant u \leqslant 1$, $v = 0$ of the real axis, traversed twice. It follows that the domain $\mathscr{G}$ consists of all points of the w-plane except those belonging to the segment Γ, i.e., $\mathscr{G} = \Gamma^c$, where Γ^c is the complement of Γ.

To study the mapping (10.40) in more detail, we find the images of the circles $|z| = r$ and the rays $\operatorname{Arg} z = \alpha + 2k\pi$ (see Figure 10.10), confining ourselves to the domain $G = I(\gamma)$. Setting

$$z = re^{it} \qquad (0 \leqslant t \leqslant 2\pi),$$

where $0 < r < 1$, we find that

$$w = u + iv = \frac{1}{2}\left(re^{it} + \frac{1}{r}e^{-it}\right) = \frac{1}{2}\left(\frac{1}{r} + r\right)\cos t - i\frac{1}{2}\left(\frac{1}{r} - r\right)\sin t,$$

or

$$u = \frac{1}{2}\left(\frac{1}{r} + r\right)\cos t, \qquad v = -\frac{1}{2}\left(\frac{1}{r} - r\right)\sin t, \tag{10.43}$$

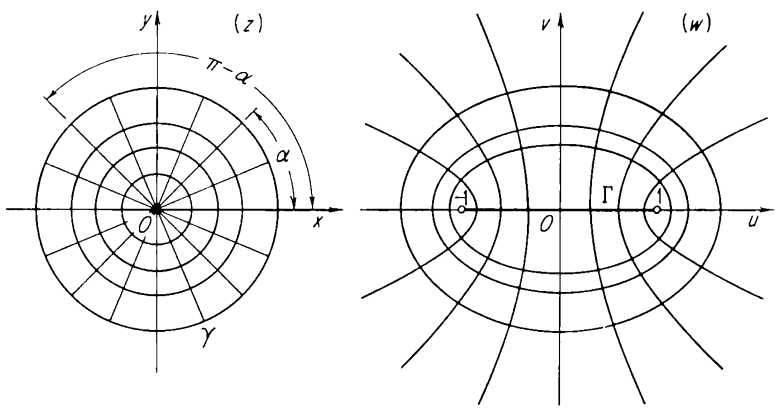

FIGURE 10.10

where $0 \leqslant t \leqslant 2\pi$. Eliminating t from (10.43), we obtain

$$\frac{u^2}{\left[\frac{1}{2}\left(\frac{1}{r} + r\right)\right]^2} + \frac{v^2}{\left[\frac{1}{2}\left(\frac{1}{r} - r\right)\right]^2} = 1, \tag{10.44}$$

which is the equation of an ellipse with semiaxes $a = \frac{1}{2}(r^{-1} + r)$, $b = \frac{1}{2}(r^{-1} - r)$ and foci at ± 1. It follows from (10.43) that as t increases continuously from 0 to 2π, i.e., as the point z describes the whole circle $|z| = r$ once in the counterclockwise direction, the image point w describes the whole ellipse once in the clockwise direction. By varying the radius of the circle $|z| = r$ from 0 to 1, we cause a to decrease from ∞ to 1 and b to decrease from ∞ to 0; as a result, the ellipses (10.44) range over the whole set of ellipses with foci ± 1. It follows, without recourse to the general considerations of Chap. 6, that $w = \lambda(z)$ is a one-to-one mapping of the unit disk $G = I(\gamma)$ onto the domain $\mathscr{G} = \Gamma^c$.

Next we consider the images of the ray

$$z = te^{i\alpha} \qquad (0 \leqslant t < 1) \tag{10.45}$$

with inclination α.[28] Substituting (10.45) into (10.40), we obtain

$$w = u + iv = \frac{1}{2}\left(\frac{1}{t} + t\right)\cos\alpha - i\frac{1}{2}\left(\frac{1}{t} - t\right)\sin\alpha$$

or

$$u = \frac{1}{2}\left(\frac{1}{t} + t\right)\cos\alpha, \qquad v = -\frac{1}{2}\left(\frac{1}{t} - t\right)\sin\alpha \qquad (0 \leqslant t < 1). \tag{10.46}$$

It follows that the images of two radii symmetric with respect to the real axis (i.e., such that if one radius has inclination α, the other has inclination $-\alpha$) are themselves symmetric with respect to the real axis, while the images of two radii symmetric with respect to the imaginary axis (i.e., such that if one radius has inclination α, the other has inclination $\pi - \alpha$) are themselves symmetric with respect to the imaginary axis. Thus it is sufficient to consider the images of radii lying in the first quadrant $0 \leqslant \alpha \leqslant \pi/2$.

For $\alpha = 0$ we have

$$u = \frac{1}{2}\left(\frac{1}{t} + t\right), \quad v = 0 \qquad (0 \leqslant t < 1),$$

which represents the infinite interval $1 < u \leqslant +\infty$. The infinite interval $-\infty \leqslant u < -1$ is the image of the radius with inclination π. For $\alpha = \pi/2$ we have

$$u = 0, \quad v = -\frac{1}{2}\left(\frac{1}{t} - t\right) \qquad (0 \leqslant t < 1),$$

which represents the negative imaginary axis $-\infty \leqslant v < 0$. The positive imaginary axis $0 < v \leqslant +\infty$ is the image of the radius with inclination $\alpha = -\pi/2$. Thus the image of the "horizontal" diameter of the unit disk $G: |z| < 1$ is the infinite interval of the real axis going from -1 to $+1$ through the point at infinity and excluding the points ± 1, while the image of the "vertical" diameter of G is the whole imaginary axis including the point at infinity but excluding the origin.

Suppose now that $0 < \alpha < \pi/2$. Then, eliminating the parameter t from (10.46), we obtain

$$\frac{u^2}{\cos^2\alpha} - \frac{v^2}{\sin^2\alpha} = 1,$$

which is the equation of a hyperbola, which we denote by H, with semiaxes $a = \cos\alpha$, $b = \sin\alpha$ and foci at ± 1. Let H_1, H_2, H_3 and H_4 denote the intersections of H with the first, second, third and fourth quadrants,

[28] By the *inclination* of a ray (or directed line segment), we mean the angle measured from the positive real axis to the ray. The slope of the ray equals the tangent of its inclination.

respectively, excluding the two points $(\pm a, 0)$ of H belonging to the real axis. Moreover, let R_α, $R_{\pi-\alpha}$, $R_{\pi+\alpha}$ and $R_{-\alpha}$ denote the sets of points belonging to the radii (10.45), with inclinations α, $\pi - \alpha$, $\pi + \alpha$, and $-\alpha$, respectively. Then it is easy to see that

$$H_1 = \lambda(R_{-\alpha}), \quad H_2 = \lambda(R_{\pi+\alpha}), \quad H_3 = \lambda(R_{\pi-\alpha}), \quad H_4 = \lambda(R_\alpha).$$

In particular, the image of each of the diameters $R_\alpha \cup R_{\pi+\alpha}$ and $R_{\pi-\alpha} \cup R_{-\alpha}$ of G is a set consisting of two "quarter-branches" of the hyperbola H, joined at infinity and minus the points $(\pm a, 0)$.

To summarize, *the function*

$$w = \lambda(z) = \frac{1}{2}\left(z + \frac{1}{z}\right)$$

is a one-to-one continuous mapping of both the interior and the exterior of the unit circle γ onto the complement of the segment $-1 \leqslant u \leqslant +1$ of the real axis. Under this mapping, the one-parameter family of circles $|z| = r$ $(0 < r < 1)$ is transformed into a one-parameter family of confocal ellipses, with semiaxes $\frac{1}{2}(r^{-1} \pm r)$ and foci at ± 1, and the one-parameter family of pairs of diameters of γ symmetric with respect to the real axis, formed of the radii

$$z = \pm te^{\pm i\alpha} \qquad (0 \leqslant t < 1),$$

where $0 < \alpha < \pi/2$,[29] *is transformed into a one-parameter family of confocal hyperbolas (minus their vertices), with semiaxes $\cos\alpha$, $\sin\alpha$ and foci at ± 1.*

Remark. Since the derivative

$$\lambda'(z) = \frac{1}{2}\left(1 - \frac{1}{z^2}\right)$$

is nonzero for $z \neq 1$, the mapping is conformal at all points of the domains $I(\gamma)$ and $E(\gamma)$. It follows that the hyperbolas intersect the ellipses at the same angles as the radii intersect the circles, i.e., at right angles. A similar situation has already been encountered in Sec. 41.

We now study the images under $w = \lambda(z)$ of circles passing through the points $z = \pm 1$. It follows from

$$w = \frac{1}{2}\left(z + \frac{1}{z}\right) = \frac{z^2 + 1}{2z} \tag{10.47}$$

that

$$w - 1 = \frac{z^2 - 2z + 1}{2z} = \frac{(z-1)^2}{2z},$$

$$w + 1 = \frac{z^2 + 2z + 1}{2z} = \frac{(z+1)^2}{2z},$$

and hence

$$\frac{w-1}{w+1} = \left(\frac{z-1}{z+1}\right)^2. \tag{10.48}$$

[29] The cases $\alpha = 0$ and $\alpha = \pi/2$ warrant special discussion (see above).

Moreover, it is easy to see that (10.48) implies (10.47). Therefore we see that the mapping $w = \lambda(z)$ is the result of making the following three mappings in succession:

$$\tilde{z} = \frac{z-1}{z+1}, \quad \tilde{w} = \tilde{z}^2, \quad w = \frac{1+\tilde{w}}{1-\tilde{w}}. \tag{10.49}$$

The first of these mappings carries any circle γ (not necessarily the unit circle) passing through the points ± 1 into a straight line passing through the origin, the second mapping carries the straight line into a ray emanating from the origin, and the third mapping carries the ray into a circular arc δ joining the points ± 1. From (10.49) we also see that if the angle between γ and the positive real axis at the point $z = 1$ equals θ, then the angle between its image δ and the positive real axis at the point $w = 1$ equals 2θ (see Figure 10.11). Moreover, it is clear that each of the two arcs of γ with end points ± 1 is separately mapped into the same arc δ.

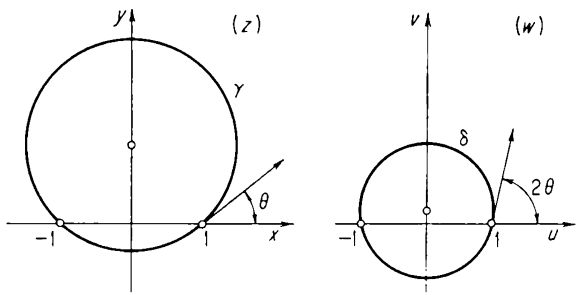

FIGURE 10.11

Remark 1. The first of the mappings (10.49) transforms the exterior of the circle γ into a half-plane, the second mapping transforms the half-plane into a domain whose boundary is a ray emanating from the origin, and the third transforms this domain into a domain whose boundary is the arc δ. Since all three mappings are one-to-one and conformal on the appropriate domains, the function $w = \lambda(z)$ is a one-to-one conformal mapping of the exterior of the circle γ (and also of its interior) onto a domain whose boundary is the circular arc δ joining the points ± 1. This generalizes the case where γ is the unit circle, and δ is the line segment joining -1 to $+1$.

Remark 2. Let Π_U be the upper half-plane $\operatorname{Im} z > 0$, let Π_L be the lower half-plane $\operatorname{Im} z < 0$, and let K be the unit disk $|z| < 1$. Then the function $w = \lambda(z)$ maps $K_U = K \cap \Pi_U$ onto Π_L and $K_L = K \cap \Pi_L$ onto Π_U. Moreover

$$\lambda(\Pi_U - \bar{K}_U) = \lambda(K_L) = \Pi_U,$$

since $\lambda(z) = \lambda(1/z)$. The image of the semicircle γ separating $\Pi_U - \bar{K}_U$ from K_U is the segment $\Gamma: -1 \leqslant u \leqslant +1$, traversed just once. Therefore the

image of the upper half-plane under $w = \lambda(z)$ is the whole plane except for the infinite segment of the real axis which joins -1 to $+1$ and passes through the point at infinity, with K_U going into Π_L and $\Pi_U - \bar{K}_U$ going into Π_U, as shown in Figure 10.12.

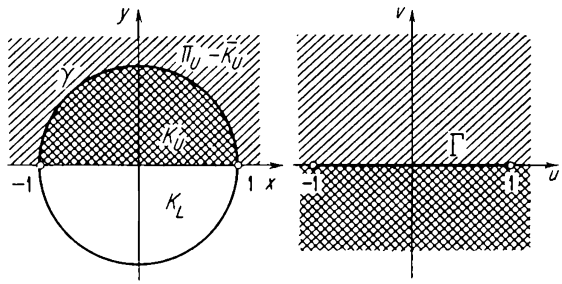

FIGURE 10.12

52. Transcendental Meromorphic Functions. Trigonometric Functions

In this section, we study the simplest *transcendental meromorphic functions*, by which we mean meromorphic functions which are not rational functions. For example, the meromorphic functions

$$\tan z = \frac{\sin z}{\cos z}, \quad \cot z = \frac{\cos z}{\sin z}, \quad \sec z = \frac{1}{\cos z}, \quad \csc z = \frac{1}{\sin z} \tag{10.50}$$

are all transcendental, since, unlike rational functions, they have infinitely many poles (points where the functions become infinite). The functions (10.50), as well as the entire functions $\cos z$ and $\sin z$, belong to the class of *trigonometric functions*, i.e., meromorphic functions of the form[29]

$$f(z) = \frac{P(e^{iz})}{Q(e^{iz})} = \frac{a_0 + a_1 e^{iz} + \cdots + a_m e^{imz}}{b_0 + b_1 e^{iz} + \cdots + b_n e^{inz}}. \tag{10.51}$$

Thus a trigonometric function is a rational function of the complex exponential e^{iz}. Obviously, we obtain the same class by considering functions which can be represented in the form

$$F(z) = \frac{\sum_{j=-m}^{m} a_j e^{ijz}}{\sum_{k=-n}^{n} b_k e^{ikz}}.$$

[29] In Sec. 40 we studied the two particularly simple *entire* trigonometric functions $\cos z$ and $\sin z$.

This last expression can be written as

$$F(z) = \frac{a_0 + \sum_{j=1}^{m} [(a_j + a_{-j}) \cos jz + i(a_j - a_{-j}) \sin jz]}{b_0 + \sum_{k=1}^{n} [(b_k + b_{-k}) \cos kz + i(b_k - b_{-k}) \sin kz]}$$

$$= \frac{a_0 + \sum_{j=1}^{m} (A'_j \cos jz + A''_j \sin jz)}{b_0 + \sum_{k=1}^{n} (B'_k \cos kz + B''_k \sin kz)}.$$

In the special case where the denominator is a constant, we can set $b = 1$, obtaining a *trigonometric polynomial*

$$F(z) = a_0 + \sum_{j=1}^{m} (A'_j \cos jz + A''_j \sin jz).$$

If at least one of the numbers A'_m and A''_m is nonzero, $F(z)$ is said to be a trigonometric polynomial of *degree m*.

Obviously, every trigonometric function is periodic with period 2π. Therefore it is sufficient to study its behavior in any strip $G: x_0 \leqslant x < x_0 + 2\pi$ parallel to the imaginary axis (note that G is neither open nor closed), since the function behaves in exactly the same way in all the strips

$$G_k: x_0 + 2k\pi \leqslant x < x_0 + (2k + 1)\pi,$$

where $k = 0, \pm 1, \pm 2, \ldots$ and $G_0 = G$. As z ranges over the strip G (including the line $x = x_0$), the variable $z_1 = iz$ ranges over a strip $x_0 \leqslant y < x_0 + 2\pi$ of the same width 2π parallel to the real axis, and hence $t = e^{iz}$ describes an angle of 2π radians with its vertex at the origin. The sides of this angle coalesce into a single ray $\operatorname{Arg} t = x_0 + 2k\pi$, which is swept out as z moves along the line $\operatorname{Re} z = x_0$. Therefore, as z ranges over the whole strip G, $t = e^{iz}$ ranges over the whole plane, taking every value except $t = 0$ and $t = \infty$ (recall that e^z does not approach a limit as $z \to \infty$).

Returning to the function (10.51), we write

$$R(t) = \frac{P(t)}{Q(t)},$$

so that

$$f(z) = R(e^{iz}).$$

Suppose the polynomials $P(t)$ and $Q(t)$ are of degrees m and n, respectively, so that the rational function $R(t)$ is of order $N = \max(m, n)$. Then, as we know from Sec. 43, for every complex number A, the equation

$$R(t) = A \tag{10.52}$$

has N roots in the extended t-plane, and there are at most $m + n$ values of A for which (10.52) can have multiple roots. However, in our case, the values $t = 0$ and $t = \infty$ are excluded, since they are not possible values of the function $t = e^{iz}$. Therefore we can only assert that (10.52) has N roots (some of which may be multiple) for every value of A with the possible exception of the numbers $R(0)$ and $R(\infty)$. Since the correspondence between z and t is one-to-one on the strip G, and since the behavior of the function $z = e^{it}$ is identical on all the strips G_k, where $k = 0, \pm 1, \pm 2, \ldots$ and $G_0 = G$, we have proved the following result:

THEOREM 10.12. *If*

$$f(z) = R(e^{iz}) \not\equiv \text{const}$$

is a trigonometric function, the equation

$$f(z) = A$$

has infinitely many roots for every complex number A with the possible exception of $R(0)$ and $R(\infty)$, for which the equation may have no roots at all.

Example 1. Let

$$f(z) = e^{inz},$$

so that

$$R(t) = t^n,$$

where $t = e^{iz}$. Then $R(0) = 0$, $R(\infty) = \infty$, and moreover the function $R(t)$ does not vanish or become infinite unless $t = 0$ or $t = \infty$. Therefore $f(z)$ does not vanish or become infinite for any z.

Example 2. Let

$$f(z) = \sec z = \frac{2e^{iz}}{1 + e^{2iz}},$$

so that

$$R(t) = \frac{2t}{1 + t^2}.$$

Then $R(0) = R(\infty) = 0$, and moreover the function $R(t)$ does not vanish unless $t = 0$ or $t = \infty$. Therefore $f(z)$ does not vanish for any value of z.

Example 3. Let

$$f(z) = \tan z = \frac{1}{i}\frac{e^{2iz} - 1}{e^{2iz} + 1},$$

so that

$$R(t) = \frac{1}{i}\frac{t^2 - 1}{t^2 + 1}.$$

Since $R(0) = i$ and $R(\infty) = -i$, and since $R(t)$ does not equal $\pm i$ unless $t = 0$ or $t = \infty$, the function $f(z)$ does not equal $\pm i$ for any z.

Example 4. Let

$$f(z) = \frac{\cos z}{\sin 2z} = i\frac{e^{3iz} + e^{iz}}{e^{4iz} - 1},$$

so that

$$R(t) = i\frac{t^3 + t}{t^4 - 1}$$

and $R(0) = R(\infty) = 0$. However, $R(t)$ vanishes at points other than 0 and ∞, i.e., for $t = \pm i$. Therefore, in this case, $f(z)$ takes *all* complex values at infinitely many points.

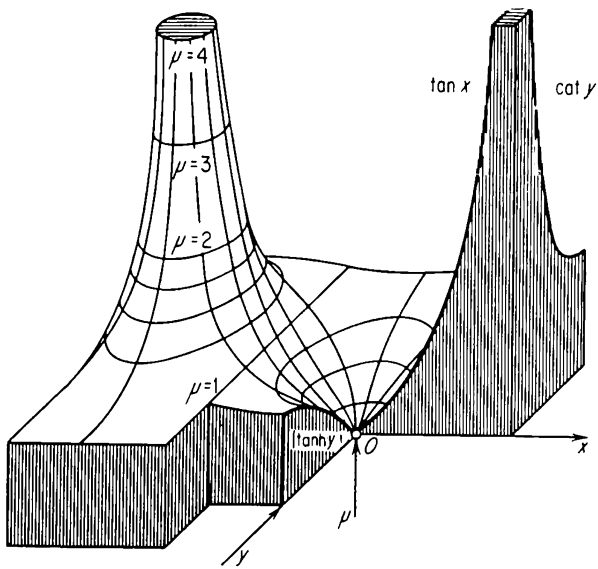

FIGURE 10.13

We now study the function

$$f(z) = \tan z = \frac{1}{i}\frac{t^2 - 1}{t^2 + 1} \qquad (t = e^{iz})$$

in more detail. In Figure 10.13 we plot the *modular surface* of $\tan z$, i.e., the surface $\mu = |\tan z|$.[30] The mapping $w = \tan z$ can be regarded as the result of making the following four mappings in succession:

$$\zeta = iz, \quad t = e^{\zeta}, \quad \tau = t^2, \quad w = \frac{1}{i}\frac{\tau - 1}{\tau + 1}. \tag{10.53}$$

[30] The unlabelled curves correspond to the equation $\arg \tan z = \text{const}$.

If G is the strip

$$x_0 < x < x_0 + h \qquad (0 < h \leqslant \pi) \tag{10.54}$$

parallel to the real axis, the mapping $\zeta = iz = \xi + i\eta$ carries G into the strip

$$x_0 < \eta < x_0 + h \tag{10.55}$$

parallel to the imaginary axis. Then the mapping $t = e^{\zeta}$ carries the strip (10.55) into the interior of the angle of h radians with vertex at the origin and sides

$$\operatorname{Arg} t = x_0 + 2k\pi, \quad \operatorname{Arg} t = x_0 + h + 2l\pi \qquad (k, l = 0, \pm 1, \pm 2, \ldots).$$

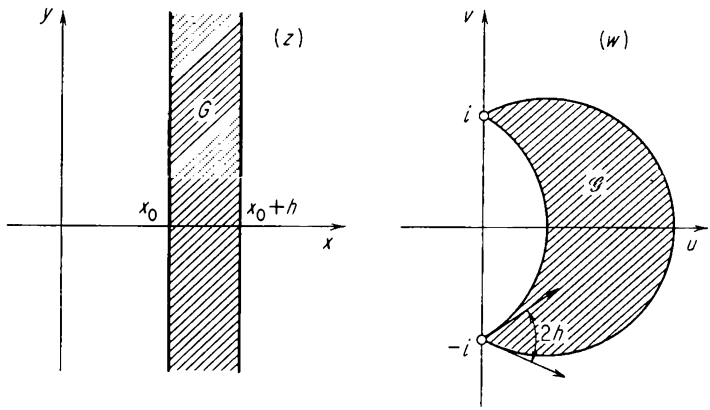

FIGURE 10.14

Under the mapping $\tau = t^2$, this angle is carried into the domain A which is the interior of the angle of $2h$ radians with sides

$$\operatorname{Arg} \tau = 2x_0 + 2k\pi, \qquad \operatorname{Arg} \tau = 2x_0 + 2h + 2l\pi, \tag{10.56}$$

and finally, the Möbius transformation

$$w = \frac{1}{i}\frac{\tau - 1}{\tau + 1} \tag{10.57}$$

carries the rays (10.56) into circular arcs joining the points $w = i$ ($\tau = 0$) and $w = -i$ ($\tau = \infty$), while carrying A itself into a circular lune with angles of $2h$ radians and vertices $\pm i$ (see Figure 10.14). Since

$$\left.\frac{dw}{d\tau}\right|_{\tau=0} = -2i, \qquad \operatorname{Arg}(-2i) = -\frac{\pi}{2} + 2k\pi,$$

the mapping (10.57) rotates the tangent to any curve emanating from the point $\tau = 0$ through the angle $-\frac{1}{2}\pi$. Therefore, at the point $w = i$, the tangents to the circular arcs bounding the lune must have inclinations $2x_0 - \frac{1}{2}\pi + 2k\pi$ and $2x_0 + 2h - \frac{1}{2}\pi + 2l\pi$. These conditions completely determine the domain $\mathcal{G} = f(G)$.

Remark. Each of the mappings (10.53) is one-to-one and conformal on the appropriate domains. It follows that $w = \tan z$ is a one-to-one conformal mapping of the strip (10.54) onto a circular lune with angles $2h$ and vertices $\pm i$. In particular, if $h = \pi/2$, the angles of the lune become π, and the lune itself becomes a circle.

PROBLEMS

10.1 Given a rational function

$$f(z) = \frac{P(z)}{Q(z)}, \tag{10.58}$$

prove that the mapping $w = f(z)$ is conformal at any simple zero of $Q(z)$, and also at $z = \infty$ if the equation $f(z) = f(\infty)$ has no multiple roots.

10.2. Prove that the mapping (10.58) actually fails to be conformal at each of the points $\gamma_1, \ldots, \gamma_r$ which are the roots of the equation

$$P'(z)Q(z) - P(z)Q'(z) = 0$$

[see (10.8)], and also at the point $\gamma_0 = \infty$ if the equation $f(z) = f(\infty)$ has multiple roots. In particular, prove that an angle with vertex at γ_j is enlarged a number of times equal to the multiplicity of the root γ_j of the equation $f(z) = f(\gamma_j)$, $j = 0, 1, \ldots, r$.

10.3. Prove that the Möbius transformations of the special form

$$\begin{aligned} &L_1 = z, \qquad L_2 = \frac{1}{z}, \qquad L_3 = 1 - z, \\ &L_4 = \frac{1}{1 - z}, \quad L_5 = \frac{z - 1}{z}, \quad L_6 = \frac{z}{z - 1} \end{aligned} \tag{10.59}$$

form a group.

Comment. This fact is summarized by saying that the transformations (10.59) are a *subgroup* of $\mathcal{M}$, the group of Möbius transformations (see footnote 13, p. 184).

10.4. Find the images of the following domains under the indicated Möbius transformations:

a) The quadrant $x > 0$, $y > 0$ if $w = \dfrac{z - i}{z + i}$;

b) The half-disk $|z| < 1$, $\operatorname{Im} z > 0$ if $w = \dfrac{2z - i}{2 + iz}$;

c) The sector $0 < \arg z < \dfrac{\pi}{4}$ if $w = \dfrac{z}{z - 1}$;

d) The strip $0 < x < 1$ if $w = \dfrac{z - 1}{z}$ or if $w = \dfrac{z - 1}{z - 2}$.

10.5. Prove that every Möbius transformation $w = L(z)$ with a single finite fixed point z_0 satisfies a relation of the canonical form

$$\frac{1}{w - z_0} = \frac{1}{z - z_0} + h \qquad (h \neq 0).$$

10.6. Prove that every Möbius transformation $w = L(z)$ with two distinct finite fixed points z_1 and z_2 satisfies a relation of the canonical form

$$\frac{w - z_1}{w - z_2} = k \frac{z - z_1}{z - z_2}.$$

10.7. Suppose the Möbius transformation $L(z)$ has two distinct finite fixed points z_1 and z_2. Prove that

$$L'(z_1)L'(z_2) = 1.$$

Hint. Use the preceding problem.

10.8. Find the Möbius transformation which carries the points -1, i, $1 + i$ into the points

a) 0, $2i$, $1 - i$; b) i, ∞, 1.

Ans. a) $w = -\dfrac{2i(z + 1)}{4z - 1 - 5i}$.

10.9. Find the Möbius transformation which carries the points -1, ∞, i into the points

a) i, 1, $1 + i$; b) ∞, i, 1; c) 0, ∞, 1.

Ans. b) $w = \dfrac{iz + 2 + i}{z + 1}$.

10.10. Find the Möbius transformation such that

a) The points 1 and i are fixed, but the point 0 goes into the point -1;
b) The points $\frac{1}{2}$ and 2 are fixed, but the point $\frac{5}{4} + \frac{3}{4}i$ goes into ∞;
c) The point i is the only fixed point and the point 1 goes into ∞.

Ans. b) $w = \dfrac{(1 - 4i)z - 2(1 - i)}{2(1 - i)z - (4 - i)}$.

10.11. Find the general form of a Möbius transformation which maps the upper half-plane $\operatorname{Im} z > 0$ onto

a) Itself; b) The lower half-plane; c) The right half-plane.

10.12. Find the Möbius transformation $w = L(z)$ mapping the upper half-plane $\operatorname{Im} z > 0$ onto itself, and satisfying the conditions

a) $w(0) = 1$, $w(1) = 2$, $w(2) = \infty$;
b) $w(0) = 1$, $w(i) = 2i$;
c) $w(a) = b$, $\arg w'(a) = \alpha$ ($\operatorname{Im} a > 0$, $\operatorname{Im} b > 0$).

Ans. b) $w = -2\left(\dfrac{2z + 1}{z - 2}\right)$.

10.13. Find the point symmetric to the point $2 + i$ with respect to
a) The circle $|z| = 1$; b) The circle $|z - i| = 3$.

10.14. Invert each of the following curves in the unit circle:
a) $|z| = \frac{1}{2}$; b) $|z - 1| = 1$; c) $y = 2$; d) $|z - z_0| = |z_0|$;
e) $|z - z_0| = \sqrt{|z_0|^2 - 1}$ $(|z_0| > 1)$; f) $x^2 - y^2 = 1$;
g) The boundary of the triangle with nonzero vertices z_1, z_2, z_3.

10.15. Map the upper half-plane $\operatorname{Im} z > 0$ onto the unit disk $|w| < 1$ in such a way that
a) $w(i) = 0$, $\arg w'(i) = -\pi/2$;
b) $w(2i) = 0$, $\arg w'(2i) = 0$;
c) $w(\alpha) = 0$, $\arg w'(\alpha) = 0$ $(\operatorname{Im} \alpha > 0)$.

Ans. b) $w = i\dfrac{z - 2i}{z + 2i}$.

10.16. Map the disk $|z| < 2$ onto the half-plane $\operatorname{Re} w > 0$ in such a way that $w(0) = 1$, $\arg w'(0) = \pi/2$.

10.17. Map the disk $|z - 4i| < 2$ onto the half-plane $v > u$ in such a way that the center of the disk goes into the point -4, while the point $2i$ on the circumference of the disk goes into the origin.

10.18. In formula (10.35), the argument of μ is left unspecified. If the problem is to have a unique solution, we must impose an extra condition on $L(z)$. Find the form of the mapping (10.35) if
a) A given point a on the circle $C: |z| = R$ goes into the point $w = R$ on C;
b) The derivative $L'(\alpha)$ is a positive real number.

Comment. Cf. the corresponding conditions on p. 182 for the mapping (10.33).

10.19. Find the Möbius transformation $w = L(z)$ mapping the unit disk $|z| < 1$ onto itself in such a way that
a) $w(\frac{1}{2}) = 0$, $\arg w'(\frac{1}{2}) = 0$; b) $w(\frac{1}{2}i) = 0$, $\arg w'(\frac{1}{2}i) = \pi/2$;
c) $w(0) = 0$, $\arg w'(0) = -\pi/2$; d) $w(\alpha) = \alpha$, $\arg w'(\alpha) = 0$ $(|\alpha| < 1)$.

10.20. Find the Möbius transformation mapping the unit disk $|z| < 1$ onto itself in such a way that two given points z_1 and z_2 of the disk go into the points $\pm\alpha$ $(|\alpha| < 1)$. Find the expression for α in terms of z_1 and z_2.

10.21. Consider the Möbius transformation

$$w = e^{i\lambda}\frac{z - \alpha}{1 - \bar{\alpha}z} \qquad (|\alpha| < 1), \tag{10.60}$$

which maps the disk $K: |z| \leqslant 1$ onto itself. Find the isometric circle (see Sec. 33) of the transformation (10.60), and identify the part of K where the magnification is less than 1. Find the maximum and minimum of $|dw/dz|$ on K.

10.22. Prove that the transformation (10.60) has two distinct fixed points unless $|\alpha| = \sin(\lambda/2)$, in which case it has only one fixed point.

10.23. Given two points z_1 and z_2 of the disk $K: |z| < R$ representing the $\mathscr{L}$-plane, find the $\mathscr{L}$-distance $\lambda(z_1, z_2)$ between z_1 and z_2, i.e., the $\mathscr{L}$-length of the $\mathscr{L}$-segment joining z_1 and z_2.

Ans. $\lambda(z_1, z_2) = \ln \dfrac{R + r}{R - r}$, where $r = \left| \dfrac{z_2 - z_1}{R^2 - \bar{z}_1 z_2} \right|$.

10.24. Let z_1, z_2 and z_3 be three points of the $\mathscr{L}$-plane. Prove that $\mathscr{L}$-distance satisfies the "triangle inequality"

$$\lambda(z_1, z_3) \leqslant \lambda(z_1, z_2) + \lambda(z_2, z_3),$$

where equality holds if and only if the point z_3 lies on the $\mathscr{L}$-segment joining z_1 to z_3.

10.25. What is the smallest possible value of the sum of the angles of an $\mathscr{L}$-triangle?

10.26. By an *$\mathscr{L}$-circle* is meant the locus of points in the $\mathscr{L}$-plane whose $\mathscr{L}$-distance from a given point is a constant. Which curves of the disk $K: |z| < R$ representing the $\mathscr{L}$-plane correspond to $\mathscr{L}$-circles?

10.27. Construct another model of Lobachevskian geometry based on Euclidean concepts in the upper half-plane $\Pi_U: \operatorname{Im} z > 0$ instead of in the disk $K: |z| < R$. Interpret all the concepts of Sec. 50 in this new model.

Hint. Use a Möbius transformation to map K onto Π_U.

10.28. Find the domain onto which the Joukowski function

$$w = \lambda(z) = \frac{1}{2}\left(z + \frac{1}{z}\right)$$

maps the disk $K: |z| < 1$ *cut along* the segment $[a, 1]$, $-1 < a < 1$, i.e., with the segment $[a, 1]$ deleted.

10.29. Let G be the exterior of the unit circle $|z| = 1$. What are the inverse images under the Joukowski function of the families of ellipses and hyperbolas in the w-plane described in Sec. 51?

10.30. What is the inverse image in the z-plane of the unit disk $|w| < 1$ under the mapping defined by the equation

$$w + \frac{1}{w} = \frac{4a}{(1 + a)^2}\left(z + \frac{1}{z}\right) - 2\frac{(1 - a)^2}{(1 + a)^2} \qquad (0 < a < 1)?$$

Ans. The disk $|z| < 1$ cut along the segment $[a, 1]$.

10.31. Study the mapping of the strip $0 < x < \pi$ under the function $w = \cos z$ by regarding $w = \cos z$ as the result of the following sequence of mappings:

$$\tilde{z} = iz, \quad \tilde{w} = \exp \tilde{z} \quad w = \frac{1}{2}\left(\tilde{w} + \frac{1}{\tilde{w}}\right).$$

10.32. Which of the following functions takes all complex values including ∞:

a) $f(z) = \dfrac{\cos 3z}{\sin z}$; b) $f(z) = \cot 3z$?

10.33. Find the images of the following strips under the mapping $w = \tan z$:

a) The strip $0 < x < \pi$; b) The strip $0 < x < \pi/4$;

c) The half-strip $0 < x < \pi$, $y > 0$.

10.34. What family of strips is mapped into the interior of the unit circle $|w| = 1$ by the function $w = \tan z$?

10.35. Find the images of the following strips under the mapping $w = \coth z = \cosh z/\sinh z$:

a) The strip $0 < y < \pi$; b) The half-strip $0 < y < \pi$, $x > 0$.

11

ELEMENTARY MULTIPLE-VALUED FUNCTIONS

53. Single-Valued Branches. Univalent Functions

Except for the Möbius transformations, which are one-to-one mappings of the extended plane onto itself, the entire and meromorphic functions $w = f(z)$ studied so far are such that the equation $f(z) = A$ has multiple roots, which are distinct except for certain special values of A. This means that the mapping $w = f(z)$ is not one-to-one, i.e., that the inverse function $z = f^{-1}(w)$ is multiple-valued. Before the concepts and results obtained for single-valued functions can be applied to the multiple-valued function $f^{-1}(w)$, we must find domains on which $f^{-1}(w)$ is no longer multiple-valued, thereby constructing so-called *single-valued branches* of $f^{-1}(w)$. This has already been done in certain special cases, e.g., for the inverses of the functions

$$w = (z - a)^n, \quad w = e^z, \quad w = \cos z$$

(see Secs. 37, 39, 42).[1] The general procedure goes as follows:

Suppose $w = f(z)$ is a single-valued function which is defined and wide-sense continuous on a domain G of the extended z-plane, but which is not one-to-one on G. Suppose we can find a countable family of disjoint subdomains $G_1 \subset G$, $G_2 \subset G, \ldots$ such that every point of G is either a point of one of the subdomains $G_1, G_2, \ldots$ or a common boundary point of at least two subdomains G_k, G_l, and such that the function $w = f(z)$ is one-to-one

[1] See also the preliminary discussion of the concept of a single-valued branch in Sec. 30.

on every subdomain G_k. Thus, if E is the set of all points of G which are common boundary points of at least two subdomains G_k, G_l, we have the decomposition

$$G = E \cup G_1 \cup G_2 \cup \cdots. \tag{11.1}$$

Then every image $\mathcal{G}_1 = f(G_1)$, $\mathcal{G}_2 = f(G_2), \ldots$ is also a domain (see Theorem 6.1), and

$$\mathcal{G} = f(G) = f(E) \cup \mathcal{G}_1 \cup \mathcal{G}_2 \cup \cdots.$$

By hypothesis, the function $z = f^{-1}(w)$ is multiple-valued on $\mathcal{G}$, i.e., $f^{-1}(w)$ can take any of a whole set of values $E_w \subset G$ at any point $w \in \mathcal{G}$. However, suppose we now define on each domain $\mathcal{G}_k$ a function $f_k^{-1}(w)$ such that

$$f_k^{-1}(\{w\}) = E_w \cap G_k \qquad (w \in \mathcal{G}_k).$$

Since the set $E_w \cap G_k$ consists of one and only one point, $f_k^{-1}(w)$ is single-valued and (wide-sense) continuous on $\mathcal{G}_k$ (see Theorem 6.1), and obviously $G_k = f_k^{-1}(\mathcal{G}_k)$. Each of the functions $f_k^{-1}(w)$, $k = 1, 2, \ldots$, of which there may be infinitely many, is called a *single-valued branch* of the function $f^{-1}(w)$.

Remark 1. It should be noted that the character of the domains $\mathcal{G}_k$ and of the single-valued branches $f_k^{-1}(w)$ depends in an essential way on just how the domain G is decomposed into subdomains G_k. In the simplest cases, a decomposition of G can be found such that all the domains $\mathcal{G}_k$ are the same.

Remark 2. For an arbitrary wide-sense continuous function $w = f(z)$, the decomposition (11.1) is not possible. However, considerations which will not be given here show that if the function $f(z) \not\equiv$ const is wide-sense continuous on a domain G, and analytic on G except possibly on a set $I \subset G$ consisting entirely of isolated points (see Problem 3.16), then the decomposition (11.1) is always possible (actually, in infinitely many ways). A function $f(z)$ which is wide-sense continuous on a domain G, and analytic on G except possibly on a set $I \subset G$ consisting entirely of isolated points, is said to be *univalent* (synonymously, *schlicht* or *simple*) on G if $f(z_1) \neq f(z_2)$ whenever $z_1, z_2 \in G$ and $z_1 \neq z_2$, i.e., if $f(z)$ is one-to-one on G. On the other hand, a function $f(z)$ which is wide-sense continuous on a domain G, and analytic on G except possibly on a set $I \subset G$ consisting entirely of isolated points, is said to be *multivalent* on G if there exists at least one pair of points $z_1, z_2 \in G$, $z_1 \neq z_2$, such that $f(z_1) = f(z_2)$. With this terminology, the result just mentioned takes the following form: If the function $f(z) \not\equiv$ const is multivalent on a domain G, then G has a decomposition (11.1) such that $f(z)$ is univalent on every subdomain G_k. The domains G_k $(k = 1, 2, \ldots)$ are called *domains of univalence* for the function $w = f(z)$. Moreover, the inverse function $z = f^{-1}(w)$ is single-valued, in fact univalent, on each of the domains $\mathcal{G}_k = f(G_k)$, $k = 1, 2, \ldots$ (cf. Rule 5, p. 109).

In this chapter, we shall illustrate the above considerations by applying them to certain elementary multiple-valued functions. However, we shall not have to rely on the result cited in Remark 2, since in every case the decomposition of the domain G into domains of univalence can be obtained by using known properties of elementary functions.

54. The Mapping $w = \sqrt[n]{z}$

Let $n > 1$ be an integer, and consider the function

$$w = \sqrt[n]{z}, \tag{11.2}$$

which is the inverse of the function $z = w^n$. For every value of z except 0 and ∞, (11.2) takes n different values, given by the formula

$$w = \sqrt[n]{|z|}\left(\cos\frac{\operatorname{Arg} z}{n} + i\sin\frac{\operatorname{Arg} z}{n}\right). \tag{11.3}$$

For $z = 0$ or $z = \infty$, the function $w = \sqrt[n]{z}$ takes just one value, i.e., $w = 0$ or $w = \infty$. The n numbers (11.3), representing the points of the w-plane at which w^n takes the same value z, correspond to the vertices of a regular n-gon, inscribed in the circle $|w| = \sqrt[n]{|z|}$. Conversely, the vertices of any regular n-gon with center at the origin of coordinates represent n possible values of the function (11.2), for a suitable complex number z. Therefore, a domain in the w-plane will be a domain of univalence for the function $z = w^n$ if and only if it contains no more than one vertex of every regular n-gon with center $w = 0$. Obviously, this condition is satisfied by the interior of every angle of $2\pi/n$ radians with vertex at $w = 0$.

As already noted, the inverse of (11.2) is the multivalent function $z = w^n$, defined on the whole w-plane. Suppose we draw any n rays from the point $w = 0$ such that the angles between adjacent rays all equal $2\pi/n$. Then the interiors $\mathcal{G}_1, \ldots, \mathcal{G}_n$ of the n angles of $2\pi/n$ radians formed by these rays are all domains of univalence for the function $z = w^n$. The image (under $z = w^n$) of each of these domains $\mathcal{G}_k$ is the same domain G in the z-plane, whose boundary is some ray drawn from the point $z = 0$. In fact, if the boundary of $\mathcal{G}_k$ consists of the rays with slopes

$$\varphi_0 + \frac{2k\pi}{n} \quad \text{and} \quad \varphi_0 + \frac{2(k+1)\pi}{n},$$

the boundary of G consists of the single ray L with slope $n\varphi_0$. In this way, we obtain n single-valued branches

$$(\sqrt[n]{z})_1, \ldots, (\sqrt[n]{z})_n \tag{11.4}$$

of the function $\sqrt[n]{z}$, all defined on the same domain G, where $(\sqrt[n]{z})_k$ denotes

the branch which maps G onto $\mathscr{G}_k$. Moreover, since $w = (\sqrt[n]{z})_k$ is a one-to-one continuous mapping of G onto $\mathscr{G}_k$, and since $z = w^n$ has a nonzero derivative nw^{n-1} on $\mathscr{G}_k$, the branches $(\sqrt[n]{z})_k$ all have nonzero derivatives on G, i.e.,

$$\frac{d}{dz}(\sqrt[n]{z})_k = \frac{1}{nw^{n-1}} = \frac{1}{n(\sqrt[n]{z})_k^{n-1}} \qquad (k = 1, \ldots, n)$$

(cf. Rule 5, p. 109).

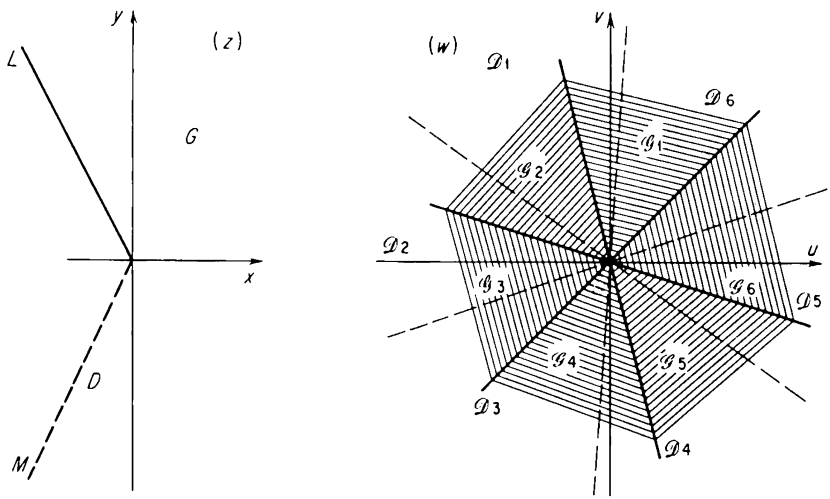

FIGURE 11.1

Now suppose we rotate our family of n rays through an angle α about the origin, where $0 < \alpha < 2\pi/n$, thereby obtaining a new family of rays, which divides the w-plane into a new family of domains $\mathscr{D}_1, \ldots, \mathscr{D}_n$. Each domain $\mathscr{D}_k$ intersects two domains $\mathscr{G}_k$ and $\mathscr{G}_{k+1}$, with $\mathscr{G}_{n+1} = \mathscr{G}_1$ by definition (see Figure 11.1 illustrating the case $n = 6$, where boundaries of the domains $\mathscr{G}_k$ are indicated by solid lines, and boundaries of the domains $\mathscr{D}_k$ by dashed lines). The inverse image in the z-plane of each of the domains $\mathscr{D}_k$ is the same domain D, whose boundary is the single ray M drawn from the origin with inclination $n\varphi_0 + n\alpha$. As before, we can define n single-valued branches [2]

$$\{\sqrt[n]{z}\}_1, \ldots, \{\sqrt[n]{z}\}_n \tag{11.5}$$

of the function $w = \sqrt[n]{z}$, where now $\{\sqrt[n]{z}\}_k$ is the branch mapping D onto

[2] Note the vital distinction between the parentheses in (11.4) and the braces in (11.5).

$\mathscr{D}_k$. Again, each of the branches (11.5) is differentiable on D, and

$$\frac{d}{dz}\{\sqrt[n]{z}\}_k = \frac{1}{n\{\sqrt[n]{z}\}_k^{n-1}} \qquad (k = 1, \ldots, n).$$

Moreover, it is clear that $\{\sqrt[n]{z}\}_k$ coincides with $(\sqrt[n]{z})_k$ on the set $\mathscr{D}_k \cap \mathscr{G}_k$ and with $(\sqrt[n]{z})_{k+1}$ on the set $\mathscr{D}_k \cap \mathscr{G}_{k+1}$. Thus, when we go from one family of domains of univalence to another such family, each new single-valued branch is obtained by combining two of the old single-valued branches, where on the part of $\mathscr{D}_k$ belonging to the common boundary Γ of $\mathscr{G}_k$ and $\mathscr{G}_{k+1}$, $\{\sqrt[n]{z}\}_k$ is the appropriate limit of either $(\sqrt[n]{z})_k$ or $(\sqrt[n]{z})_{k+1}$. More precisely, we have

$$\begin{aligned}
\{\sqrt[n]{z}\}_k &= (\sqrt[n]{z})_k \qquad \text{if} \quad z \in \mathscr{D}_k \cap \mathscr{G}_k, \\
\{\sqrt[n]{z}\}_k &= (\sqrt[n]{z})_{k+1} \qquad \text{if} \quad z \in \mathscr{D}_k \cap \mathscr{G}_{k+1}, \\
\{\sqrt[n]{z}\}_k &= \lim_{\zeta \to z} (\sqrt[n]{\zeta})_k = \lim_{\zeta \to z} (\sqrt[n]{\zeta})_{k+1} \qquad \text{if} \quad z \in \mathscr{D}_k \cap \Gamma,
\end{aligned}$$

where $\Gamma = \bar{\mathscr{G}}_k \cap \bar{\mathscr{G}}_{k+1}$.

Remark. If the angle of rotation α is zero, then $D = G$ and $\mathscr{D}_k = \mathscr{G}_k$ $(k = 1, \ldots, n)$, while if $\alpha = 2\pi/n$, $D = G$ again but $\mathscr{D}_k = \mathscr{G}_{k+1}$ $(k = 1, \ldots, n)$, where $\mathscr{G}_{n+1} = \mathscr{G}_1$. As α increases continuously from 0 to $2\pi/n$, the domain $\mathscr{D}_k$ overlaps the domain $\mathscr{G}_{k+1}$ more and more, until it finally coincides with $\mathscr{G}_{k+1}$, and the ray M representing the boundary of D undergoes a counterclockwise rotation of 2π radians, where its initial and final positions coincide with the ray L representing the boundary of G. At the same time, the branch $\{\sqrt[n]{z}\}_k$, which originally coincides with $(\sqrt[n]{z})_k$, shares more and more of its domain of definition with the branch $(\sqrt[n]{z})_{k+1}$, until it finally coincides with $(\sqrt[n]{z})_{k+1}$. In this sense, we can say that as α increases continuously from 0 to $2\pi/n$, the branch $(\sqrt[n]{z})_k$ changes continuously into the branch $(\sqrt[n]{z})_{k+1}$.

We can also keep track of the way one branch $(\sqrt[n]{z})_k$ changes into another branch $(\sqrt[n]{z})_{k+1}$ by making the point z describe a complete circle with center at the point $z = 0$. Suppose that at the point $z_0 \in G$ we choose a value of $\sqrt[n]{z}$ belonging to the branch $(\sqrt[n]{z})_k$ and represented by the point

$$w_0 = \sqrt[n]{|z_0|}\left(\cos\frac{\theta_0}{n} + i\sin\frac{\theta_0}{n}\right),$$

belonging to the domain $\mathscr{G}_k$. Then, as the point z moves continuously around the circle $|z| = |z_0|$ in the counterclockwise direction, starting from the point z_0, the value of

$$w = \sqrt[n]{|z_0|}\left(\cos\frac{0}{n} + i\sin\frac{0}{n}\right) \tag{11.6}$$

varies continuously with 0, and when z returns to its original value z_0, (11.6) goes into the value

$$w_1 = \sqrt[n]{|z_0|}\left(\cos\frac{\theta_0 + 2\pi}{n} + i\sin\frac{\theta_0 + 2\pi}{n}\right),$$

obtained by rotating w_0 through the angle $2\pi/n$ about the point $w = 0$.[3] Therefore w_1 belongs to the domain $\mathscr{G}_{k+1}$ adjacent to $\mathscr{G}_k$, and w_1 is the value of the branch $(\sqrt[n]{z})_{k+1}$ at the point z_0. Since the point $z_0 \in G$ is arbitrary, we can say that one circuit around the origin $z = 0$ in the counterclockwise direction causes the branch $(\sqrt[n]{z})_k$ to change continuously into the branch $(\sqrt[n]{z})_{k+1}$.[4] Moreover, it is easy to see that in this sense n circuits around the origin in the counterclockwise sense cause the branch $(\sqrt[n]{z})_k$ to undergo the sequence of transformations

$$(\sqrt[n]{z})_k \to (\sqrt[n]{z})_{k+1},\ (\sqrt[n]{z})_{k+1} \to (\sqrt[n]{z})_{k+2},\ \ldots,$$
$$(\sqrt[n]{z})_n \to (\sqrt[n]{z})_1,\ \ldots,\ (\sqrt[n]{z})_{k-1} \to (\sqrt[n]{z})_k,$$

which carry it continuously into itself after "going through" all the other branches in succession. Since $(\sqrt[n]{z})_k$ is arbitrary, n circuits around the origin carry any branch into itself.

Given a multiple-valued function $w = f(z)$ with continuous single-valued branches defined on a domain G, we say that the point $\zeta \in \bar{G}$ is a *branch point* of $f(z)$ if there exists a neighborhood $\mathscr{N}(\zeta)$ such that one complete circuit around an arbitrary closed Jordan curve $\gamma \subset \mathscr{N}(\zeta)$ with $\zeta \in I(\gamma)$, carries every branch of $f(z)$ into another branch of $f(z)$. If a finite number of circuits around γ (in the same direction) carries every branch of $f(z)$ into itself, and if n is the smallest such number, we say that ζ is a branch point *of finite order*, specifically, *of order* $n - 1$. In this case, the point ζ is also called an *algebraic branch point* of $f(z)$, provided that $f(z)$ has a limit (finite or infinite) at ζ. Thus we have just shown that the point $z = 0$ is an algebraic branch point of order $n - 1$ of the function $w = \sqrt[n]{z}$.

Remark 1. It is clear that the point $z = \infty$ can also be regarded as an algebraic branch point of order $n - 1$ of the function $w = \sqrt[n]{z}$, since every circuit around the point at infinity along a circle of arbitrarily large radius with center at the origin is simultaneously a circuit around the origin. Therefore the multiple-valued function $w = \sqrt[n]{z}$ has two branch points in the z-plane, i.e., $z = 0$ and $z = \infty$, both of order $n - 1$.

Remark 2. The single-valued branches described above were constructed

---

[3] Of course, in making the circuit around the circle $|z| = |z_0|$, we allow z to pass through the ray L, which is excluded from the domain G.

[4] More precisely, every value of $\sqrt[n]{z}$ on the branch $(\sqrt[n]{z})_k$ changes continuously into the corresponding value of $\sqrt[n]{z}$ on the branch $(\sqrt[n]{z})_{k+1}$.

for a domain like G or D, whose boundary is a rectilinear ray joining the two branch points 0 and ∞. More generally, let γ be any Jordan curve in the extended z-plane joining the points 0 and ∞, and this time let G be the domain with boundary γ. As the point z traces out the curve γ from its initial point 0 to its final point ∞, the n points corresponding to the n values of $w = \sqrt[n]{z}$ trace out n Jordan curves $\Gamma_1, \ldots, \Gamma_n$ joining 0 and ∞. These curves have no points in common other than 0 and ∞, and each set $\Gamma_k \cup \Gamma_{k+1}$ (where $\Gamma_{k+1} = \Gamma_1$) represents a closed Jordan curve in the extended w-plane. Of the two domains with boundary $\Gamma_k \cup \Gamma_{k+1}$, let $\mathcal{G}_k$ be the domain which does not contain the other curves $\Gamma_1, \ldots, \Gamma_{k-1}, \Gamma_{k+2}, \ldots, \Gamma_n$. By construction, when the w-plane is rotated through the angle $2\pi/n$ about the origin, Γ_k goes into Γ_{k+1} and Γ_{k+1} goes into Γ_{k+2}, and hence the domain $\mathcal{G}_k$ goes into the domain $\mathcal{G}_{k+1}$ ($\mathcal{G}_{n+1} = \mathcal{G}_1$). Since

$$\mathcal{G}_k \cap \mathcal{G}_{k+1} = 0 \qquad (k = 1, \ldots, n),$$

the rotation cannot carry any point of $\mathcal{G}_k$ into another point of $\mathcal{G}_k$. Therefore the domains $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are all domains of univalence for the function $w = z^n$, and we obtain n single-valued branches of the function $w = \sqrt[n]{z}$, all defined on the domain G, by requiring that the kth branch take its values in the domain $\mathcal{G}_k$ ($k = 1, \ldots, n$). To specify a branch, it is sufficient to indicate the value of $\sqrt[n]{z}$ at some point $z_0 \in G$; if this value is w_0, there is a unique domain $\mathcal{G}_k$ containing w_0, and a unique branch of $\sqrt[n]{z}$ taking the value w_0 at the point z_0.

Now let $[\sqrt[n]{z}]_k$ and $[\sqrt[n]{z}]_l$ be two single-valued branches of the function $\sqrt[n]{z}$, which are defined on the domain G and take values w_0' and w_0'', respectively, at a point $z_0 \in G$. Since

$$w_0' = [\sqrt[n]{z}]_k = \sqrt[n]{|z|}\left(\cos\frac{\theta_0 + 2m'\pi}{n} + i\sin\frac{\theta_0 + 2m'\pi}{n}\right),$$
$$w_0'' = [\sqrt[n]{z}]_l = \sqrt[n]{|z|}\left(\cos\frac{\theta_0 + 2m''\pi}{n} + i\sin\frac{\theta_0 + 2m''\pi}{n}\right),$$

where $\theta_0 = \arg z$, and m', m'' are integers, it follows that w_0'' equals w_0' multiplied by

$$\varepsilon_l = \cos\frac{2(m'' - m')\pi}{n} + i\sin\frac{2(m'' - m')\pi}{n},$$

i.e., by a value of $\sqrt[n]{1}$. But $\varepsilon_l[\sqrt[n]{z}]_k$ is obviously a single-valued continuous function on G, such that $(\varepsilon_l[\sqrt[n]{z}]_k)^n = z$, i.e., $\varepsilon_l[\sqrt[n]{z}]_k$ is one of the single-valued branches of $\sqrt[n]{z}$ defined on G, in fact, the branch $[\sqrt[n]{z}]_l$ containing the point $\varepsilon_l[\sqrt[n]{z_0}]_k = [\sqrt[n]{z_0}]_l = w_0''$. In other words, any single-valued branch of $\sqrt[n]{z}$ defined on G can be obtained by multiplying any other single-valued branch defined on G by an appropriate nth root of unity.

Remark 3. The conclusions of this section apply (with certain obvious modifications) to the somewhat more general functions

$$w = \sqrt[n]{z - a} \quad \text{and} \quad w = \sqrt[n]{\frac{z - a}{z - b}}, \tag{11.7}$$

which are the inverses of the functions

$$z = w^n + a \quad \text{and} \quad z = \frac{bw^n - a}{w^n - 1},$$

respectively. The first of the functions (11.7) has branch points a and ∞, while the second has branch points a and b. Moreover, single-valued branches of each of the functions (11.7) can be defined on any domain whose boundary is a Jordan curve joining the appropriate branch points.

55. The Mapping $w = \sqrt[n]{P(z)}$

To gain a deeper insight into the concept of a branch point, we now study the multiple-valued function

$$w = f(z) = \sqrt[n]{P(z)},$$

where $P(z)$ is an arbitrary polynomial of degree N. Let $P(z)$ have zeros $a_1, \ldots, a_m$, of orders $\alpha_1, \ldots, \alpha_m$, respectively, where $\alpha_1 + \cdots + \alpha_m = N$. Then, according to Sec. 35, $P(z)$ can be written in the form

$$P(z) = A(z - a_1)^{\alpha_1} \cdots (z - a_m)^{\alpha_m},$$

and hence

$$f(z) = \sqrt[n]{A(z - a_1)^{\alpha_1} \cdots (z - a_m)^{\alpha_m}}. \tag{11.8}$$

Consider an arbitrary closed Jordan curve γ which does not pass through any of the points $a_1, \ldots, a_m$, and suppose z traverses γ once. At some point $z_0 \in \gamma$ we choose definite values $\varphi_1^{(0)}, \ldots, \varphi_m^{(0)}$ of the arguments of the complex numbers $z_0 - a_1, \ldots, z_0 - a_m$, thereby selecting a certain single-valued branch of the function $f(z)$.

As the point z goes around the curve γ once, starting from z_0 and returning to z_0, the argument φ_k of the vector $z - a_k$ varies continuously; if a_k belongs to $E(\gamma)$ [the exterior of γ], φ_k returns to its original value $\varphi_k^{(0)}$, while if a_k belongs to $I(\gamma)$ [the interior of γ], φ_k acquires an increment $\pm 2\pi$ (see Figure 11.2).[5] The sign of the increment depends only on the direction

[5] These facts, which are easily verified in the simplest cases (for example, when γ is a circle, an ellipse or a polygon), can be proved in complete generality. See e.g., P. S. Aleksandrov, *op. cit.*, Chap. 2.

in which γ is traversed, and if the appropriate angles *increase*, we say that the direction is *positive* (this is always the counterclockwise direction). Suppose, for definiteness, that the point z describes γ in the positive direction. Then, if none of the points $a_1, \ldots, a_m$ lie inside γ, all the angles φ_k return to their original values, and as a result, the function $f(z)$ also returns to its original value (11.8). It follows that *a point ζ of the finite z-plane different from the zeros $a_1, \ldots, a_m$ of the polynomial $P(z)$ cannot be a branch point of $\sqrt[n]{P(z)}$*. In fact, for any such point ζ we can find a neighborhood $\mathcal{N}(\zeta)$ containing none of the points $a_1, \ldots, a_m$, and then a complete circuit around any closed Jordan curve $\gamma \subset \mathcal{N}(\zeta)$ with $\zeta \in I(\gamma)$ does not change the branch of $f(z)$ which has been chosen.

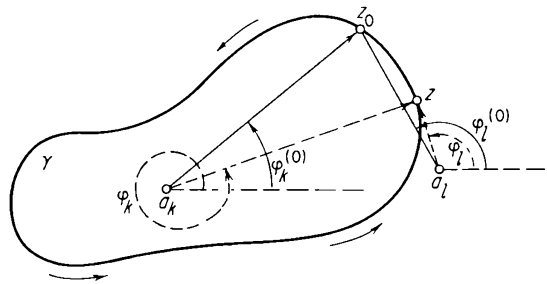

FIGURE 11.2

Next let $\mathcal{N}(a_k)$ be a neighborhood of the point a_k which is small enough not to contain any of the other points $a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_m$. Then a complete circuit around any closed Jordan curve $\gamma \subset \mathcal{N}(a_k)$ with $a_k \in I(\gamma)$ changes φ_k by 2π, while all the other angles $\varphi_1, \ldots, \varphi_{k-1}, \varphi_{k+1}, \ldots, \varphi_n$ return to their original values. It follows that the right-hand side of (11.8) is multiplied by the factor

$$\cos \frac{2\pi\alpha_k}{n} + i \sin \frac{2\pi\alpha_k}{n}, \tag{11.9}$$

which is different from unity if and only if α_k is not a multiple of n. Therefore *every zero a_k of the polynomial $P(z)$ whose order is not a multiple of n is a branch point of $\sqrt[n]{P(z)}$*. To determine the order of such a branch point, suppose δ_k is the greatest common divisor of α_k and n ($\delta_k < n$). Then, setting $\alpha_k = \delta_k \alpha_k'$ and $n = \delta_k \nu_k$ ($\nu_k > 1$), we see that (11.9) equals

$$\cos \frac{2\pi\alpha_k'}{\nu_k} + i \sin \frac{2\pi\alpha_k'}{\nu_k}.$$

Therefore, as a result of p circuits around γ (in the same direction), $f(z)$ is multiplied by the factor

$$\cos \frac{2\pi\alpha_k' p}{\nu_k} + i \sin \frac{2\pi\alpha_k' p}{\nu_k},$$

which is equal to unity if and only if p is a multiple of ν_k. Since ν_k is obviously the smallest multiple of ν_k, *the branch point* a_k *is of order* $\nu_k - 1$.

Finally, let $\mathcal{N}(\infty)$ be a neighborhood of the point at infinity which contains none of the points $a_1, \ldots, a_m$, and let γ be a closed Jordan curve such that $\gamma \subset \mathcal{N}(\infty)$ and $\infty \in E(\gamma)$. Since $I(\gamma)$ contains all the points $a_1, \ldots, a_m$, a complete circuit around γ in the positive direction changes all the angles $\varphi_1, \ldots, \varphi_m$ by 2π. Therefore the right-hand side of (11.8) is multiplied by the factor

$$\cos \frac{2\pi(\alpha_1 + \cdots + \alpha_m)}{n} + i \sin \frac{2\pi(\alpha_1 + \cdots + \alpha_m)}{n} = \cos \frac{2\pi N}{n} + i \sin \frac{2\pi N}{n},$$

which is different from unity if and only if N is not a multiple of n. Therefore ∞ *is a branch point of* $\sqrt[n]{P(z)}$ *if and only if* N *is not a multiple of* n. Suppose N is not divisible by n, so that ∞ is a branch point of $\sqrt[n]{P(z)}$. Let δ be the greatest common divisor of N and n ($\delta < n$), and let $n = \delta\nu$. Then *the branch point* ∞ *is of order* $\nu - 1$.

Remark. Let γ be any closed Jordan curve lying in the finite z-plane. As we have just seen, a circuit around γ does not change the values of $f(z) = \sqrt[n]{P(z)}$ if either of the following two conditions is met:

1. $a_k \in I(\gamma)$, $a_j \in E(\gamma)$ for $j \neq k$, and α_k is a multiple of n;
2. $a_k \in I(\gamma)$ for $k = 1, \ldots, m$, and $N = \alpha_1 + \cdots + \alpha_m$ is a multiple of n.

More generally, let $a_{k_1}, \ldots, a_{k_q}$ ($q \leqslant m$) be any set of zeros of $P(z)$, such that $\alpha_{k_1} + \cdots + \alpha_{k_q}$ is a multiple of n. Then a circuit around any closed Jordan curve γ, such that $I(\gamma)$ contains $a_{k_1}, \ldots, a_{k_q}$ and $E(\gamma)$ contains all the other zeros, does not change the values of $f(z) = \sqrt[n]{P(z)}$.

Now let G be a domain such that every closed Jordan curve γ lying in G has the property that either $I(\gamma)$ contains no zeros of $P(z)$ at all, or else $I(\gamma)$ contains a set of zeros the sum of whose orders is divisible by n. Then on every such domain G we can define single-valued branches of the function $f(z) = \sqrt[n]{P(z)}$. In fact, let $z_0 \in G$ and let w_0 be one of the n values of the function $f(z)$ at z_0. The single-valued branch $f(z)$ which takes the value w_0 at z_0 is constructed as follows: To find the value of this branch at any other point $z_1 \in G$ we draw a Jordan curve $L \subset G$ joining z_0 and z_1 (see Theorem 4.12), and we move along z_0 to z_1, making sure that the corresponding values of $f(z)$ vary continuously, starting from the initial value w_0 at z_0. As a result, we arrive at z_1 "accompanied" by one of the n values of $f(z)$, which we denote by w_1. It remains to show that w_1 is unique, i.e., that w_1 depends

only on z_0, w_0 and z_1, and not on the particular Jordan curve joining z_0 to z_1. Suppose that by going from z_0 to z_1 along another Jordan curve $L' \subset G$ we arrive at z_1 accompanied by a value w_1' of $f(z)$, where $w_1' \neq w_1$. Without loss of generality, we can assume that L and L' have only the points z_0 and z_1 in common (see Problem 11.3). Then $\gamma = L \cup L'$ is a closed Jordan curve such that $\gamma \subset G$ and $z_1 \in \gamma$, but such that one circuit around γ beginning and ending at z_1 changes the value of $f(z)$ at z_1 from w_1 to w_1'. But this is impossible, since by hypothesis γ either contains no zeros of $P(z)$ or a set of zeros the sum of whose orders is divisible by n. This contradiction establishes the uniqueness of w_1.

Example 1. The function

$$w = f(z) = \sqrt{(1 - z^2)(1 - k^2z^2)} \qquad (0 < k < 1) \tag{11.10}$$

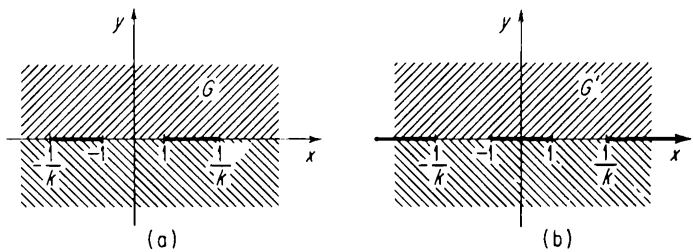

FIGURE 11.3

is a double-valued function with four branch points ± 1, $\pm 1/k$. Here $N = 4$ is a multiple of $n = 2$, and hence ∞ is not a branch point. Since ± 1, $\pm 1/k$ are all simple zeros of the expression under the radical, the numbers α_k all equal 1. Therefore a circuit around any closed Jordan curve γ containing only two branch points in its interior does not change the values of the function. Thus, for example, we can define two single-valued branches of $f(z)$ on the domain G with boundary consisting of the two segments

$$-\frac{1}{k} \leqslant x \leqslant -1, \qquad 1 \leqslant x \leqslant \frac{1}{k}$$

[see Figure 11.3(a)], or on the domain G' with boundary consisting of the segment $-1 \leqslant x \leqslant 1$ and the infinite segment of the real axis joining the points $-1/k$ and $1/k$ through the point at infinity [see Figure 11.3(b)]. On the domain G, the two branches $f_1(z)$ and $f_2(z)$ of the function (11.10) can be distinguished by the values they take at the origin, i.e.,

$$f_1(0) = 1, \qquad f_2(0) = -1.$$

Example 2. Consider the function

$$w = \sqrt{4z^3 - g_2 z - g_3}, \tag{11.11}$$

where g_2 and g_3 are complex numbers satisfying the condition

$$g_2^3 - 27g_3^2 \neq 0,$$

which means that the discriminant of the cubic polynomial

$$4z^3 - g_2 z - g_3$$

is nonzero, so that the zeros e_1, e_2, e_3 of (11.11) are all distinct.[6] In this case, $N = 3$ is not divisible by $n = 2$, and hence the point at infinity is also a branch point. As before, a circuit around any pair of branch points along any closed Jordan curve does not change the value of the function. Therefore, joining e_1 to e_2 and e_2 to ∞ by Jordan curves γ_1 and γ_2, we obtain a domain G with boundary consisting of γ_1 and γ_2 on which we can define single-valued branches of the function (see Figure 11.4).

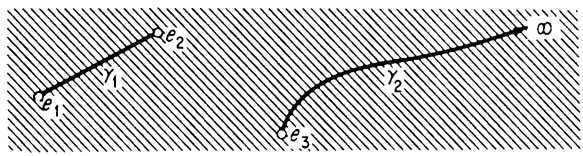

FIGURE 11.4

Example 3. Consider the function

$$w = f(z) = z + \sqrt{z^2 - 1}, \tag{11.12}$$

which is the inverse of the Joukowski function

$$z = \frac{1}{2}\left(w + \frac{1}{w}\right). \tag{11.13}$$

The function $f(z)$ is double-valued, and has the same branch points ± 1 as the function $\sqrt{z^2 - 1}$. To obtain a domain G on which single-valued branches of (11.12) can be defined, we join the points -1 and 1 by a finite segment of the real axis. As we know from Sec. 51, this gives a domain which is mapped in a one-to-one fashion by the function (11.12) onto each of the two domains $I(\gamma)$ and $E(\gamma)$, where γ is the unit circle. To

[6] See e.g., G. Birkhoff and S. MacLane, *op. cit.*, p. 113.

distinguish these branches, we can fix one of the two values of $f(z)$ at some point of the domain G, e.g., at the point ∞. According to (11.13), z becomes infinite either for $w = 0$ or $w = \infty$. Therefore the branch $f_1(z)$ of (11.12) satisfying the condition $f_1(\infty) = 0$ maps G onto $I(\gamma)$, while the branch $f_2(z)$ satisfying the condition $f_2(\infty) = \infty$ maps G onto $E(\gamma)$. Instead of G, we might have chosen the domain G' whose boundary consists of the infinite segment of the real axis joining the points -1 and 1, or the domains G_U and G_L, whose boundaries are the upper and lower unit semicircles.

56. The Logarithm

The inverse of the function

$$z = e^w = e^u(\cos v + i \sin v)$$

is defined for any value of z different from 0 and ∞, and is represented by the formula

$$w = \ln |z| + i \operatorname{Arg} z$$

[cf. (9.25)]. This function, which is obviously multiple-valued (in fact, *infinite-valued*), is called the *logarithm* and is denoted by $\operatorname{Ln} z$, i.e.,

$$\operatorname{Ln} z = \ln |z| + i \operatorname{Arg} z, \tag{11.14}$$

by definition. The value

$$\ln |z| + i \arg z$$

of the logarithm is called the *principal value*, and is denoted by $\ln z$. Then (11.14) can be written in the form

$$\operatorname{Ln} z = \ln z + 2k\pi i \qquad (k = 0, \pm 1, \pm 2, \ldots). \tag{11.15}$$

It follows that *every complex number different from* 0 *and* ∞ *has infinitely many logarithms* (*i.e., values of the function* $\operatorname{Ln} z$), *and any two of these logarithms differ by an integral multiple of* $2\pi i$.

If z is a positive real number, the principal value of the logarithm is just $\ln |z|$, which is exactly what is meant by the logarithm in elementary mathematics; for example,

$$\ln 1 = 0, \quad \ln e = 1, \quad \ln 2 = 0.69315 \ldots$$

For negative real numbers and for imaginary numbers, the principal value of the logarithm is an imaginary number

$$\ln |z| + i \arg z \qquad (\arg z \neq 0, \quad -\pi < \arg z \leqslant \pi)$$

and all the other values of the logarithm are also imaginary numbers,

calculated by using (11.15); for example,

$$\begin{aligned}
\operatorname{Ln}(-1) &= (2k+1)\pi i,\\
\operatorname{Ln}(-2) &= 0.69315\ldots + (2k+1)\pi i,\\
\operatorname{Ln}(1-i) &= \ln\sqrt{2} - \frac{\pi i}{4} + 2k\pi i = 0.34657\ldots + (8k-1)\frac{\pi i}{4}.
\end{aligned}$$

The familiar rules for finding logarithms of products and quotients remain valid for the multiple-valued logarithms of complex numbers, since

$$\begin{aligned}
\operatorname{Ln}(z_1 z_2) &= \ln|z_1 z_2| + i\operatorname{Arg}(z_1 z_2)\\
&= \ln|z_1| + \ln|z_2| + i(\operatorname{Arg} z_1 + \operatorname{Arg} z_2)\\
&= \operatorname{Ln} z_1 + \operatorname{Ln} z_2
\end{aligned} \tag{11.16}$$

and

$$\begin{aligned}
\operatorname{Ln}\frac{z_1}{z_2} &= \ln\frac{z_1}{z_2} + i\operatorname{Arg}\frac{z_1}{z_2}\\
&= \ln|z_1| - \ln|z_2| + i(\operatorname{Arg} z_1 - \operatorname{Arg} z_2)\\
&= \operatorname{Ln} z_1 - \operatorname{Ln} z_2,
\end{aligned} \tag{11.17}$$

where z_1 and z_2 are arbitrary nonzero complex numbers. In (11.16) and (11.17), both the left and right-hand sides (for fixed z_1 and z_2) represent infinite sets of complex numbers, and the equalities have to be understood in the sense that these two sets are equal, i.e., have the same members. Failure to remember this fact can lead to paradoxical results. For example, in a sophism constructed by John Bernoulli, it is claimed that $\operatorname{Ln}(-z) = \operatorname{Ln} z$ for arbitrary $z \neq 0$, and the following chain of equalities is adduced as "proof":

1. $\operatorname{Ln}[(-z)^2] = \operatorname{Ln}(z^2)$;
2. $\operatorname{Ln}(-z) + \operatorname{Ln}(-z) = \operatorname{Ln} z + \operatorname{Ln} z$;
3. $2\operatorname{Ln}(-z) = 2\operatorname{Ln} z$;
4. $\operatorname{Ln}(-z) = \operatorname{Ln} z$.

However, the conclusion that $\operatorname{Ln}(-z) = \operatorname{Ln} z$ is false, since

$$\begin{aligned}
\operatorname{Ln} z &= \ln|z| + i\operatorname{Arg} z = \ln|z| + i\arg z + 2k\pi i,\\
\operatorname{Ln}(-z) &= \ln|-z| + i\operatorname{Arg}(-z) = \ln|z| + i\arg z + (2k+1)\pi i,
\end{aligned}$$

and obviously none of the numbers representing the values of $\operatorname{Ln} z$ is the same as any of the numbers representing the values of $\operatorname{Ln}(-z)$. The fallacy in the "proof" occurs in going from equality 2 to equality 3. The first of these relations is based on formula (11.16), and is of course true. However, the sum $\operatorname{Ln}(-z) + \operatorname{Ln}(-z)$ cannot be replaced by $2\operatorname{Ln}(-z)$, since the sum in question is obtained from the set of numbers $\operatorname{Ln}(-z)$ by adding each of these numbers to itself and to *all the other* numbers of the set $\operatorname{Ln}(-z)$, whereas the set $2\operatorname{Ln}(-z)$ is obtained by simply doubling all the numbers $\operatorname{Ln}(-z)$, i.e., by adding each such number *to itself only*.[7]

[7] The following simple example may help clarify the situation: Let A be the set consisting of the two numbers 0 and 1. Then $A + A$ is the set consisting of the three numbers $0 + 0 = 0$, $0 + 1 = 1$ and $1 + 1 = 2$, whereas the set $2A$ consists only of two numbers $2 \cdot 0 = 0$ and $2 \cdot 1 = 2$.

Therefore

$$\operatorname{Ln}(-z) + \operatorname{Ln}(-z) \neq 2 \operatorname{Ln}(-z),$$

and by the same token,

$$\operatorname{Ln} z + \operatorname{Ln} z \neq 2 \operatorname{Ln} z.$$

Setting $z_1 = z_2 = z \neq 0$ in (11.17), we obtain the relation

$$\operatorname{Ln} 1 = \operatorname{Ln} z - \operatorname{Ln} z, \tag{11.18}$$

which is a correct formula. However, the right-hand side of (11.18) cannot be replaced by 0, since here we are talking about the set of all possible differences between values of the logarithm of the same number. This set consists of all possible multiples of $2\pi i$, so that to be perfectly explicit, we should write (11.18) as

$$\operatorname{Ln} 1 = 2k\pi i \qquad (k = 0, \pm 1, \pm 2, \ldots).$$

We now study the single-valued branches of the logarithm. We begin by finding domains of univalence for the exponential function $z = e^w$, which is the inverse of the logarithm $w = \operatorname{Ln} z$. All the numbers w for which e^w takes any given value z are given by the formula

$$w = \ln |z| + i \operatorname{Arg} z$$

[cf. (9.95)], i.e., all the numbers w can be obtained by shifting any one of them by $2k\pi i$, where $k = 0, \pm 1, \pm 2, \ldots$. Therefore a domain of univalence for $z = e^w$ cannot contain any pair of points such that one point can be obtained from the other by a shift of this kind. The simplest way to satisfy this requirement is to start with an open rectilinear strip

$$\mathscr{G}_0 : v_0 < v < v_0 + 2\pi$$

of width 2π parallel to the real axis in the w-plane. Then, subjecting $\mathscr{G}_0$ to all possible shifts of the form $2k\pi i$, we obtain an infinite family of domains of univalence:

$$\mathscr{G}_k : v_0 + 2k\pi < v < v_0 + 2(k+1)\pi \qquad (k = 0, \pm 1, \pm 2, \ldots).$$

Obviously, every point of the w-plane is either an interior point of a domain $\mathscr{G}_k$ or a boundary point of two domains $\mathscr{G}_k$ and $\mathscr{G}_{k+1}$ [see Figure 11.5(b)].

The image (under $z = e^w$) of each of the strips $\mathscr{G}_k$ is the same domain G in the z-plane, i.e., the interior of an angle of 2π radians with its vertex at the origin. The boundary of G is a single ray of inclination v_0 emanating from the origin [see Figure 11.5(a)]. We can define infinitely many different single-valued branches of $\operatorname{Ln} z$ on the domain G by specifying that the kth branch $\operatorname{Ln}_k z$ $(k = 0, \pm 1, \pm 2, \ldots)$ have $\mathscr{G}_k$ as its range. Each of these branches is uniquely characterized by the value w_0 which it assigns to any given

point $z_0 \in G$, since one and only one of the domains $\mathcal{G}_k$ contains the point w_0. Expressed somewhat differently,

$$\operatorname{Ln}_k z = \ln |z| + i \operatorname{Arg}_k z,$$

where $\operatorname{Arg}_k z$ is the value of the argument satisfying the inequality

$$v_0 + 2k\pi < \operatorname{Arg}_k z < v_0 + 2(k+1)\pi \qquad (k = 0, \pm 1, \pm 2, \ldots).$$

Moreover, since $w = \operatorname{Ln}_k z$ is a one-to-one continuous mapping of G onto $\mathcal{G}_k$, and since $z = e^w$ has a nonzero derivative e^w on $\mathcal{G}_k$, the branches $\operatorname{Ln}_k z$ all have nonzero derivatives on G, i.e.,

$$\frac{d}{dz} \operatorname{Ln}_k z = \frac{1}{e^w} = \frac{1}{z} \qquad (k = 0, \pm 1, \pm 2, \ldots)$$

(cf. Rule 5, p. 109).

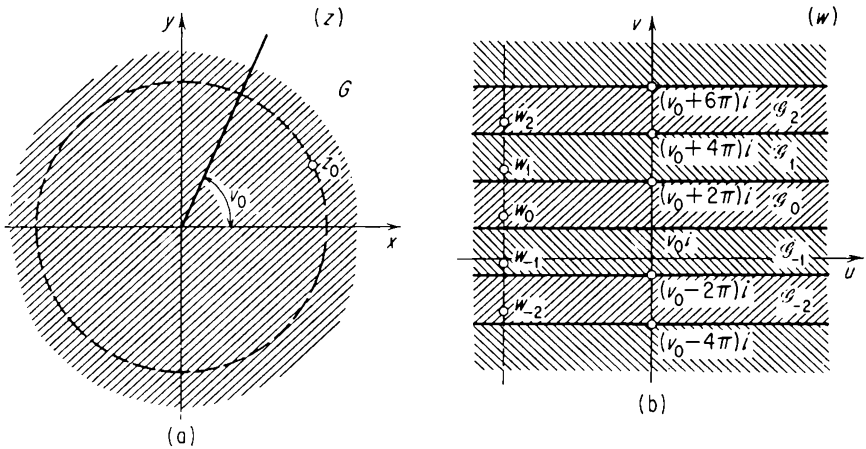

FIGURE 11.5

The points 0 and ∞ are both branch points of the function Ln z. To see this, suppose that at a point $z_0 \in G$ we choose a value of Ln z corresponding to the branch $\operatorname{Ln}_k z$ and represented by the point

$$w_k = \operatorname{Ln}_k z_0 = \ln |z_0| + i \operatorname{Arg}_k z_0,$$

belonging to the strip $\mathcal{G}_k$. Then, as z moves continuously around the circle $|z| = |z_0|$ in the counterclockwise direction, starting from the point z_0, the value of

$$w = \ln |z| + i \operatorname{Arg} z \tag{11.19}$$

changes continuously, and when z returns to its original value z_0, (11.19) goes into the value

$$w_{k+1} = \ln |z_0| + i \operatorname{Arg}_k z_0 + 2\pi i = \ln |z_0| + i \operatorname{Arg}_{k+1} z_0 = \operatorname{Ln}_{k+1} z_0.$$

Thus, since the point $z_0 \in G$ is arbitrary, one circuit around the origin $z = 0$ (or about the point at infinity $z = \infty$) causes the branch $\operatorname{Ln}_k z$ to change continuously into the branch $\operatorname{Ln}_{k+1} z$.[8] Obviously, as we make additional circuits around the origin in the counterclockwise direction, the branch $\operatorname{Ln}_k z$ undergoes the infinite sequence of transformations

$$\operatorname{Ln}_k z \to \operatorname{Ln}_{k+1} z, \quad \operatorname{Ln}_{k+1} z \to \operatorname{Ln}_{k+2} z, \quad \operatorname{Ln}_{k+2} z \to \operatorname{Ln}_{k+3} z, \ldots,$$

so that $\operatorname{Ln}_k z$ is never carried back into itself. For this reason, the points 0 and ∞ are called *branch points of infinite order* or *logarithmic branch points.*

Remark. Single-valued branches of the function $\operatorname{Ln} z$ can be defined on domains more general than G. Let γ be any Jordan curve joining the points $z = 0$ and $z = \infty$. Then the curve γ has infinitely many images Γ_k ($k = 0, \pm 1, \pm 2, \ldots$) under the mapping $w = \operatorname{Ln} z$. The curves Γ_k, which are all Jordan curves, divide the w-plane into infinitely many open curvilinear strips $\mathscr{D}_k$, where the boundary of $\mathscr{D}_k$ consists of the pair of curves Γ_k and Γ_{k+1}, and $\mathscr{D}_k$ does not contain any of the other curves Γ_k. Thus we can define a countable family of single-valued branches of $\operatorname{Ln} z$ on the domain D with boundary γ by specifying that the kth branch $(\operatorname{Ln} z)_k$, where $k = 0, \pm 1, \pm 2, \ldots$, have $\mathscr{D}_k$ as its range.[9] Clearly,

$$(\operatorname{Ln} z)_k = (\operatorname{Ln} z)_l + 2\pi(k - l)i,$$

and, as before, we find that

$$\frac{d}{dz}(\operatorname{Ln} z)_k = \frac{1}{z}. \tag{11.20}$$

Since (11.20) is independent of how the branches of $\operatorname{Ln} z$ are defined, we can simply write

$$\frac{d}{dz} \operatorname{Ln} z = \frac{1}{z},$$

where by the left-hand side we mean any single-valued branch of $\operatorname{Ln} z$, defined on a domain containing the given point z.

57. The Function z^α. Exponentials and Logarithms to an Arbitrary Base

We begin by defining the expression a^α, where $a \neq 0$ and α are arbitrary complex numbers. We first assume that α is real, and examine in turn the cases where α is an integer, a rational number and an irrational number.

---

[8] Similarly, one circuit around the origin in the opposite (i.e., clockwise) direction causes the branch $\operatorname{Ln}_k z$ to change continuously into the branch $\operatorname{Ln}_{k-1} z$.

[9] Note the distinction between $\operatorname{Ln}_k z$ and $(\operatorname{Ln} z)_k$.

Case 1. If $\alpha = n$ is an integer, then

$$a^\alpha = a^n = |a|^n [\cos (n \operatorname{Arg} a) + i \sin (n \operatorname{Arg} a)], \tag{11.21}$$

and a^α *has just one value.*

Case 2. If $\alpha = r$ is a rational number, then $r = m/n$ where m and $n > 0$ are relatively prime integers. As we already know [cf. formula (2.15)], in this case,

$$\begin{aligned} a^\alpha = a^{m/n} &= |a|^{m/n} \left[\cos \left(\frac{m}{n} \operatorname{Arg} a\right) + i \sin \left(\frac{m}{n} \operatorname{Arg} a\right)\right] \\ &= |a|^r [\cos (r \operatorname{Arg} a) + i \sin (r \operatorname{Arg} a)], \end{aligned}$$

and a^α *has n different values.*

Case 3. If $\alpha = \rho$ is an irrational number, we define a^ρ by continuity, i.e., as the limit

$$\lim_{n \to \infty} a^{r_n} = \lim_{n \to \infty} |a|^{r_n} [\cos (r_n \operatorname{Arg} a) + i \sin (r_n \operatorname{Arg} a)], \tag{11.23}$$

where $\{r_n\}$ is an arbitrary sequence of irrational numbers converging to ρ. In taking the limit (11.23), we hold Arg a fixed. Then

$$\lim_{n \to \infty} r_n \operatorname{Arg} a = \rho \operatorname{Arg} a,$$

and (11.23) implies (cf. p. 34)

$$a^\rho = |a|^\rho [\cos (\rho \operatorname{Arg} a) + i \sin (\rho \operatorname{Arg} a)]. \tag{11.24}$$

To obtain all the values of a^ρ, we now let Arg a take all its values. Since two values of ρ Arg a differ by a number of the form $2k\rho\pi$, where k is a nonzero integer, and since $k\rho$ can never be an integer, it follows that different values of Arg a give rise to different values of a^ρ. Thus, in this case, a^α *has infinitely many different values.*

Remark. It should be noted that formulas (11.21), (11.22) and (11.24) are all special cases of the formula

$$a^\alpha = |a|^\alpha [\cos (\alpha \operatorname{Arg} a) + i \sin (\alpha \operatorname{Arg} a)],$$

which can be written in the form

$$\begin{aligned} a^\alpha &= \exp (\alpha \ln |a|)[\cos (\alpha \operatorname{Arg} a) + i \sin (\alpha \operatorname{Arg} a)] \\ &= \exp (\alpha \ln |a| + i\alpha \operatorname{Arg} a) = \exp (\alpha \operatorname{Ln} a). \end{aligned} \tag{11.25}$$

For the time being, we make a distinction between $\exp z$ and e^z, with only the former being used to denote the exponential function defined in Sec. 38 (see Example 2 below).

Now let α be an arbitrary complex number. Observing that in this case the right-hand side of (11.25) still has meaning, we write

$$a^\alpha = \exp (\alpha \operatorname{Ln} a),$$

by definition. If α is imaginary, all the values of a^α corresponding to different values of Ln a (or equivalently, to different values of Arg a) are also different, since two distinct values of α Ln a differ by a number of the form $2\pi i\alpha$, which cannot be an integral multiple of $2\pi i$ if α is imaginary. It follows from (11.25) and the addition theorem for the exponential that

$$a^{\alpha_1}a^{\alpha_2} = \exp(\alpha_1 \operatorname{Ln} a)\exp(\alpha_2 \operatorname{Ln} a) = \exp[(\alpha_1 + \alpha_2)\operatorname{Ln} a] = a^{\alpha_1+\alpha_2}. \tag{11.26}$$

Moreover, we have the following rule for raising a power to a power:

$$\begin{aligned}(a^\alpha)^\beta &= [\exp(\alpha \operatorname{Ln} a)]^\beta = \exp\{\beta \operatorname{Ln}[\exp(\alpha \operatorname{Ln} a)]\} \\ &= \exp(\beta\alpha \operatorname{Ln} a) = a^{\alpha\beta}.\end{aligned} \tag{11.27}$$

Both (11.26) and (11.27) are the same as the corresponding rules for real numbers.

Example 1.

$$\begin{aligned}1^{\sqrt{2}} &= \exp(\sqrt{2}\ln 1) = \exp(2k\pi\sqrt{2}i) \\ &= \cos(2k\pi\sqrt{2}) + i\sin(2k\pi\sqrt{2}) \qquad (k = 0, \pm 1, \pm 2, \ldots).\end{aligned}$$

Example 2.

$$\begin{aligned}e^z &= \exp(z \operatorname{Ln} e) = \exp[z(1 + 2k\pi i)] \\ &= \exp z \exp(2k\pi iz) \qquad (k = 0, \pm 1, \pm 2, \ldots).\end{aligned} \tag{11.28}$$

It follows from (11.28) that only one of the values of the power e^z coincides with $\exp z$. In fact, the other values are

$$\exp z \exp(2\pi iz), \quad \exp z \exp(-2\pi iz), \ldots$$

In particular, only one of the values of e^x (where x is a real number) coincides with the positive real number $\exp x$, and in this case, the other values are

$$\exp x \exp(2\pi ix), \quad \exp x \exp(-2\pi ix), \ldots \tag{11.29}$$

[If x is rational, there are only a finite number of different values (11.29), but if x is irrational, there are infinitely many different values.] Despite this, *we shall continue to denote the exponential function by* e^z, *as well as by* $\exp z$. This use of the "multiple-valued symbol" e^z to denote a single number is analogous to the conventional use of the symbol $\sqrt[n]{a}$ (where a is a positive real number) to denote the unique positive value of the nth root of a (cf. footnote 6, p. 18).

Example 3.

$$i^i = \exp(i \operatorname{Ln} i) = \exp\left[i\left(\frac{\pi i}{2} - 2k\pi i\right)\right] = e^{(4k-1)\pi/2} \qquad (k = 0, \pm 1, \pm 2, \ldots).$$

Thus all the infinitely many values of i^i are positive real numbers, which can be arbitrarily small or arbitrarily large.

We are now in a position to study the functions z^α and a^z, where α and $a \neq 0$ are arbitrary complex numbers. First we consider the function

$$f(z) = z^\alpha$$

(which in general is defined only for $z \neq 0$), and examine in turn the cases where α is an integer, a rational number and an irrational or imaginary number.

Case 1. If $\alpha = n$ is an integer, then

$$f(z) = z^\alpha = z^n$$

is a particularly simple rational function. In this case, $f(z)$ is defined for $z = 0$, where it has a zero (if $n > 0$) or a pole (if $n < 0$).

Case 2. If $\alpha = r$ is a rational number, then $r = m/n$ where m and $n > 0$ are relatively prime integers, and hence

$$f(z) = z^{m/n} = \sqrt[n]{z^m}.$$

In this case, $f(z)$ is an n-valued function, with branch points $z = 0$ and $z = \infty$ of order $n - 1$. Let γ be any Jordan curve joining 0 and ∞, and let G be the domain with boundary γ. Then we can define n single-valued differentiable branches of $f(z)$ on G, which change continuously into each other as we go around any closed Jordan curve whose interior contains the origin.

Case 3. If α is not a rational number, i.e., if α is an irrational real number or an imaginary number, then $f(z) = z^\alpha$ is an infinite-valued function, all of whose values are given by the formula

$$z^\alpha = \exp(\alpha \operatorname{Ln} z).$$

In this case, $z = 0$ and $z = \infty$ are branch points (as in Case 2), but now they are of infinite order. In fact, if we go around the point $z = 0$ (say, in the positive direction), $\operatorname{Arg} z$ varies continuously and increases by 2π. Therefore the value of $\alpha \operatorname{Ln} z$ increases by $2\pi i\alpha$, i.e., $f(z)$ is multiplied by factor $e^{2\pi i\alpha} \neq 1$.

Next we consider the *exponential to the base a*, i.e., the function

$$a^z = \exp(z \operatorname{Ln} a),$$

defined for any finite z and any $a \neq 0$. To obtain a definite single-valued branch of $f(z)$, it is sufficient to fix one of the values of $\operatorname{Ln} a$, say

$$\operatorname{Ln} a = b = \ln a + 2k_0\pi i. \tag{11.30}$$

After this has been done, we obtain a single-valued, everywhere differentiable function $\exp(bz)$. Taking all possible values of $\operatorname{Ln} a$, we obtain all possible single-valued branches of the function a^z. Since two values of $\operatorname{Ln} a$ differ by

a term of the form $2k\pi i$, where k is an integer, two branches of the function a^z differ by a factor of the form $\exp(2k\pi iz)$. This factor is also a single-valued, everywhere differentiable function, which takes the value 1 only when z is a rational number of the form m/k (where m is an arbitrary integer).

Remark. The branches of the multiple-valued function a^z differ in an essential way from those of the multiple-valued functions considered previously. In all the cases considered so far, we can find points of the extended plane, called *branch points*, with the property that by going around each branch point along suitable closed Jordan curves we can carry any single-valued branch into any other. However, in the case of the function a^z, this is impossible, since every branch is a single-valued continuous function defined on the whole finite plane (and not on some domain whose boundary consists of certain curves joining the branch points). In fact, after making a circuit around any (finite) closed Jordan curve whatsoever, we return to the original complex number z (perhaps with a different value of the argument) and hence to the same value of the function $\exp(bz)$, where b is a fixed value of $\operatorname{Ln} a$. Thus the multiple-valued function a^z has no branch points at all, and its single-valued branches cannot be carried continuously into each other by making circuits around closed Jordan curves. In other words, the different branches can be regarded as self-contained, independent, single-valued, everywhere differentiable functions,[10] i.e., as nothing more or less than the infinite set of entire functions

$$\exp(z \ln a), \quad \exp[z(\ln a + 2\pi i)], \quad \exp[z(\ln a - 2\pi i)], \ldots$$

The fact that all these different entire functions can be represented as branches of a single infinite-valued function a^z is no more surprising than the fact that $\sin z$ and $-\sin z$ can be regarded as branches of the double-valued function

$$\sqrt{1 - \cos^2 z}, \tag{11.31}$$

or that $\cosh z$ and $\sinh z$ can be regarded as branches of the double-valued function

$$\tfrac{1}{2}[\exp z + \sqrt{\exp(-2z)}]. \tag{11.32}$$

[It should be noted that just like the function a^z, the functions (11.31) and (11.32) have no branch points.]

Finally we consider the *logarithm to the base a*, denoted by $\operatorname{Log}_a z$ and defined as the inverse of the function

$$z = a^w = \exp(w \operatorname{Ln} a). \tag{11.33}$$

---

[10] The same situation has already been encountered in Example 2 above, in connection with the function e^z (where $a = e$ and $\ln a = 1$). However, in the case of e^z, we agreed to interpret e^z as the particular single-valued branch $\exp z$, i.e., as the branch which takes real values when z is real (see p. 136). We shall make the same choice whenever a is a positive real number.

If we again write (11.30), thereby choosing a branch of a^w, (11.33) becomes

$$z = \exp(bw),$$

and hence

$$w = \operatorname{Log}_a z = \frac{1}{b} \operatorname{Ln} z, \tag{11.34}$$

which differs from Ln z only by the factor $1/b$. Since b is a value of Ln a, it follows from (11.34) that

$$\operatorname{Log}_a z = \frac{\operatorname{Ln} z}{\operatorname{Ln} a}, \tag{11.35}$$

where in the denominator we fix one of the infinitely many values of Ln a (which is then kept the same for all z). Thus, to define $\operatorname{Log}_a z$, we must specify not only the base a, but also a particular value of Ln a.

Example 1. Let $a = e$, and choose the value of Ln e equal to 1. Then

$$\operatorname{Log}_e z = \operatorname{Ln} z,$$

which is just the ordinary definition of the *natural logarithm*. However, we can also choose another value of Ln e, say $1 + 2\pi i$. In this case, (11.35) gives

$$\operatorname{Log}_e z = \frac{\operatorname{Ln} z}{1 + 2\pi i}.$$

It is easy to see that with this second definition of the natural logarithm, $\operatorname{Log}_e z$ will have a real value (in fact, exactly one) only if $z = e^k$, where k is an integer.

Example 2. Let $a = 10$, and choose the value of Ln 10 equal to

$$2.30259\ldots = \frac{1}{0.43429\ldots} = \frac{1}{M}.$$

Then we have

$$\operatorname{Log}_{10} z = M \operatorname{Ln} z = (0.43429\ldots) \operatorname{Ln} z. \tag{11.36}$$

This definition of the *common logarithm* of an arbitrary complex number $z \neq 0$ agrees with the ordinary definition of the common logarithm $\log_{10} x$ of a positive real number x. In fact, setting $z = x$ in (11.36) and taking principal values, we obtain

$$\log_{10} x = (0.43429\ldots) \ln x.$$

Example 3. If $a = 1$, we cannot define $\operatorname{Log}_1 z$ by using the principal value of Ln 1, since ln 1 = 0. Instead, suppose we choose the value of Ln 1 equal to $2\pi i$, obtaining

$$\operatorname{Log}_1 z = \frac{\operatorname{Ln} z}{2\pi i} = \frac{1}{2\pi} \operatorname{Arg} z - \frac{i}{2\pi} \ln |z|.$$

With this definition, all the values of $\operatorname{Log}_1 z$ are real if $|z| = 1$ and imaginary if $|z| \neq 1$. Thus, only numbers corresponding to points on the unit circle have real numbers as their logarithms to the base 1, and for such numbers $\operatorname{Log}_1 z = \operatorname{Arg} z$.

58. The Mapping $w = \operatorname{Arc}\cos z$

First we study the function $w = \operatorname{Arc}\cos z$, defined as the (multiple-valued) inverse of the function

$$z = \cos w.$$

Replacing $\cos w$ by

$$\frac{e^{iw} + e^{-iw}}{2},$$

and writing

$$e^{iw} = t \tag{11.37}$$

for brevity (cf. Sec. 41), we find that t satisfies the equation

$$\frac{1}{2}\left(t + \frac{1}{t}\right) = z$$

or

$$t^2 - 2zt + 1 = 0, \tag{11.38}$$

with solution[11]

$$t = z + \sqrt{z^2 - 1}. \tag{11.39}$$

It follows from (11.37) and (11.39) that

$$w = \operatorname{Arc}\cos z = -i \operatorname{Ln} t = -i \operatorname{Ln}(z + \sqrt{z^2 - 1}). \tag{11.40}$$

We now investigate the branch points of the multiple-valued function $w = \operatorname{Arc}\cos z$, which, according to (11.40), involves both a square root and a logarithm. Therefore we first examine the behavior of (11.40) at ± 1 (the branch points of $\sqrt{z^2 - 1}$), and then at $t = 0$ and $t = \infty$ (the branch points of $\operatorname{Ln} t$).

1. *The function* $\operatorname{Arc}\cos z$ *has algebraic branch points at* ± 1. In fact, as z makes one circuit around any closed Jordan curve whose interior contains either of the points ± 1 (but not the other), each value of $\sqrt{z^2 - 1}$ is replaced by its negative. Therefore each root

$$t = z + \sqrt{z^2 - 1}$$

[11] We do not write $\pm$ in front of the radical, since the square root is already understood to be double-valued [cf. (9.52)]. Note that both numbers (11.39) are nonzero (in fact, their product equals 1).

of the quadratic equation (11.38) is replaced by the other root t^{-1} (recall that the product of the two roots is 1). Correspondingly, the expression $-i \operatorname{Ln} t$ is replaced by $-i \operatorname{Ln} t^{-1}$, which differs from the first expression if $t \neq t^{-1}$. But $t = t^{-1}$ implies $t = \pm 1$ and hence $z = \pm 1$, as can be seen from (11.38) or (11.39). This case cannot occur, since by hypothesis our Jordan curve does not pass through either of the points $z = \pm 1$. Therefore the value of Arc cos z actually does change as a result of the circuit around $z = 1$ or $z = -1$, and hence the points $z = \pm 1$ are both branch points, in fact, branch points of order 1.

2. *The function* Arc cos z *has a logarithmic branch point at* ∞. According to Sec. 51, t makes a circuit around a circle in the t-plane with center $t = 0$ if and only if z makes a circuit around an ellipse in the z-plane with foci at $z = \pm 1$. Thus,as the point z traces out an ellipse with foci ± 1, Arg t changes by $\pm 2\pi$, and hence so does $-i \operatorname{Ln} t$. Since any neighborhood of the point at infinity contains an ellipse with foci ± 1, the point ∞ is a branch point of Arc cos z, in fact, a branch point of infinite order.[12]

3. *The function* Arc cos z *has no branch points in the extended z-plane except those already indicated, i.e.,* ± 1 *and* ∞. In fact, if $z_0 \neq \pm 1$, $z_0 \neq \infty$, then, corresponding to the value $z = z_0$, equation (11.38) has two roots t_0' and t_0'', which satisfy the condition $t_0' t_0'' = 1$, and are different from 0, ± 1, and ∞. We can find disjoint neighborhoods $\mathcal{N}(t_0')$ and $\mathcal{N}(t_0'')$ which are so small that they do not contain the points 0 and ∞, and are such that neither neighborhood contains a pair of points t_1, t_2 satisfying the condition $t_1 t_2 = 1$. To see this, let γ be the unit circle $|t| = 1$, and suppose that $t_0' \notin \gamma$, say $|t_0'| < 1$. Then $|t_0''| > 1$ and hence, if we choose $\mathcal{N}(t_0') \subset I(\gamma)$ and $\mathcal{N}(t_0'') \subset E(\gamma)$, neither neighborhood contains a pair of points t_1, t_2 such that $t_1 t_2 = 1$ (see Figure 11.6). On the other hand, if $t_0' \in \gamma$, then, since $t_0' \neq \pm 1$, t_0' belongs to either the upper half-plane Π_U: Im $t > 0$ or the lower half-plane Π_L: Im $t < 0$, say to Π_L. Then $t_0'' \in \gamma \cap \Pi_U$, and in this case it is sufficient to choose $\mathcal{N}(t_0') \subset \Pi_L$, $\mathcal{N}(t_0'') \subset \Pi_U$.

In any case, the function

$$z = \frac{1}{2}\left(t + \frac{1}{t}\right) \tag{11.41}$$

is one-to-one on each of the neighborhoods $\mathcal{N}(t_0')$ and $\mathcal{N}(t_0'')$, since it can only take the same value at a pair of points t_1, t_2 satisfying the

[12] A circuit around any ellipse with foci at the points $z = \pm 1$ can also be regarded as a circuit around the point $z = 0$. However, it should be noted that none of these ellipses lies entirely in the neighborhood $|z| < \varepsilon$ if $\varepsilon \leqslant 1$, and in fact it turns out that $z = 0$ is not a branch point of Arc cos z.

condition $t_1 t_2 = 1$. In other words, (11.41) is a one-to-one mapping of $\mathcal{N}(t_0')$ and $\mathcal{N}(t_0'')$ onto certain domains G' and G'' in the z-plane. Clearly, G' and G'' both contain the point z_0, which is the image of the centers of both $\mathcal{N}(t_0')$ and $\mathcal{N}(t_0'')$, and hence there is a neighborhood $\mathcal{N}(z_0)$ so small that it is contained in $G' \cap G''$. Let γ be an arbitrary closed Jordan curve such that $\gamma \subset \mathcal{N}(z_0)$ and $z_0 \in I(\gamma)$. Then the image of γ in the t-plane is a pair of closed Jordan curves Γ' and Γ'', where $\Gamma' \subset \mathcal{N}(t_0')$ and $\Gamma'' \subset \mathcal{N}(t_0'')$. As z traverses the curve γ, either t traverses Γ', corresponding to one branch of the function (11.39), or t traverses Γ'', corresponding to the other branch. However, if we make a complete circuit around Γ', starting from any point of Γ', Arg t returns to its original value, since $I(\Gamma')$ does not contain the origin of the t-plane. Exactly the same argument applies to Γ''. Therefore, a circuit around γ does not change the value of $-i \operatorname{Ln} t = \operatorname{Arc\,cos} z$, i.e., as asserted, z_0 is not a branch point of Arc cos z.

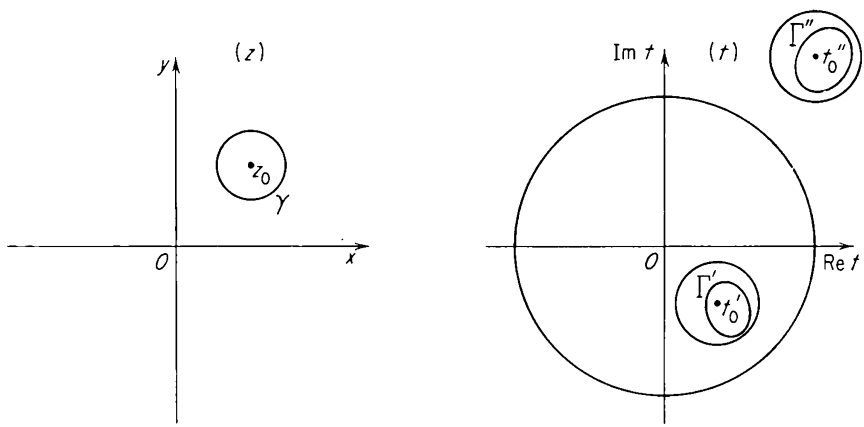

FIGURE 11.6

To find a domain G in the z-plane on which we can define single-valued continuous branches of Arc cos z, we have to connect the branch points of Arc cos z by a Jordan curve. For example, let Δ be the infinite segment of the real axis which joins -1 to $+1$ and passes through the point at infinity, and let G be the domain whose boundary is Δ. As we know from Remark 2, p. 201 the function (11.41) maps both of the half-planes Π_U: Im $t > 0$ and Π_L: Im $t < 0$ onto the domain G. Furthermore, the function

$$w = u + iv = -i \operatorname{Ln} t \tag{11.42}$$

maps both Π_U and Π_L onto certain open strips of width π parallel to the imaginary axis, i.e., (11.42) maps Π_L onto the strips

$$\mathcal{S}_{2k-1}: (2k-1)\pi < u < 2k\pi \qquad (k = 0, \pm 1, \pm 2, \ldots)$$

and Π_U onto the strips

$$\mathscr{G}_{2k}: 2k\pi < u < (2k + 1)\pi \qquad (k = 0, \pm 1, \pm 2, \ldots).$$

Therefore the strips $\mathscr{G}_n$ $(n = 0, \pm 1, \pm 2, \ldots)$ are the images of G in the w-plane.

To fix a single-valued branch of Arc cos z on the domain G, it is sufficient to specify which of the strips $\mathscr{G}_n$ is the range of the branch in question. In this way, we obtain an infinite family of single-valued branches

$$\text{Arc cos}_0\, z, \quad \text{Arc cos}_1\, z, \quad \text{Arc cos}_{-1}\, z, \ldots$$

As usual, a given single-valued branch is uniquely characterized by the value it takes at any point of G, say at $z = 0$. Of course, we can also define single-valued branches of Arc cos z on many other domains (see e.g., Problem 11.14).

59. The Mapping $w = z + \ln z$

We now study the mapping of the upper half-plane Im $z > 0$ produced by the single-valued branch

$$w = u + iv = z + \ln z \tag{11.43}$$

of the function $w = z + \text{Ln}\, z$. The simplest approach is to find the images of the rays emanating from the origin with inclinations between 0 and π. Setting $z = re^{i\theta}$ in (11.43), and separating real and imaginary parts, we can write (11.43) as a pair of equations

$$u = r\cos\theta + \ln r, \qquad v = r\sin\theta + \theta. \tag{11.44}$$

For fixed θ, these are the parametric equations of the image of the ray with inclination θ, where the parameter r varies from 0 to ∞.

First we find the image of the real axis Im $z = 0$. If $\theta = 0$, (11.44) becomes

$$u = r + \ln r, \quad v = 0,$$

and since $r + \ln r$ increases from $-\infty$ to $+\infty$ as r increases from 0 to ∞, the image of the positive real axis (including 0) in the z-plane is the whole real axis in the w-plane. If $\theta = \pi$, (11.44) becomes

$$u = -r + \ln r, \qquad v = \pi,$$

and since $-r + \ln r$ increases from $-\infty$ to -1 and then decreases from -1 to $-\infty$ as r increases from 0 to ∞, the image of the negative real axis (including 0) in the z-plane is the half-line

$$-\infty < u \leqslant -1, \qquad v = \pi \tag{11.45}$$

in the w-plane, traversed twice by the point w as the point z traverses the ray $\theta = \pi$ once.

Next we consider a ray of inclination θ, where $0 < \theta < \pi$. Since $v = r \sin \theta + \theta$, v increases from θ to $+\infty$ as r increases from 0 to ∞. Using the second of the equations (11.44) to express r in terms of v and θ, and substituting the result into the first of the equations (11.44), we obtain the equation

$$u = (v - \theta) \cot \theta + \ln \frac{v - \theta}{\sin \theta} \qquad (\theta \leqslant v < +\infty) \tag{11.46}$$

for the image of the ray of inclination θ. It is easily verified that (11.46) represents a curve with asymptote $v = \theta$, where the whole curve lies above

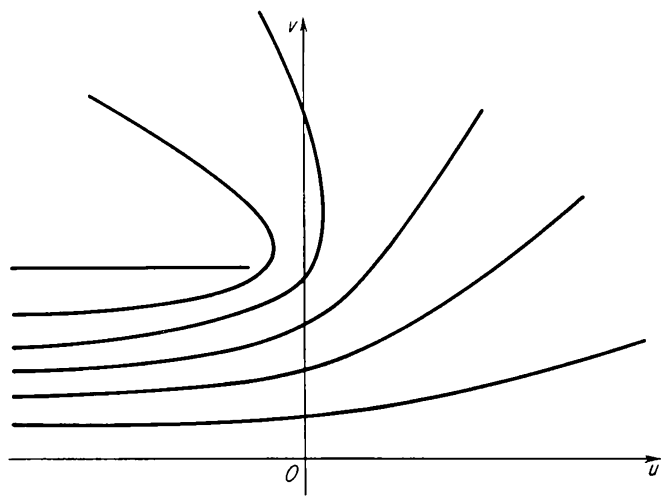

FIGURE 11.7

the asymptote. As v increases from θ to $+\infty$, u increases from $-\infty$ to $+\infty$ if $0 < \theta \leqslant \pi/2$, while if $\pi/2 < \theta < \pi$, u increases from $-\infty$ to a maximum value of

$$\ln\left(-\frac{1}{\cos \theta}\right) - 1$$

(corresponding to $v = \theta - \tan \theta$), and then decreases to $-\infty$. In every case, the curve (11.46) is "concave to the right," as shown in Figure 11.7.

Now suppose we regard u in equation (11.46) as a function of θ for fixed v $(0 \leqslant \theta \leqslant v)$. Then, since

$$\frac{\partial u}{\partial \theta} = -(v - \theta) - \frac{[1 + (v - \theta) \cot \theta]^2}{v - \theta} < 0 \qquad \text{for} \quad 0 < \theta < v,$$

u decreases from $+\infty$ to $-\infty$ as θ increases from 0 to v. Therefore two different points z' and z'' lying on two different rays $\theta = \theta'$ and $\theta = \theta''$ cannot have the same image $w_0 = u_0 + iv_0$. But it is also impossible for two points z' and z'' lying on the same ray to have the same image, since, as we have just seen, if a variable point z traces out the ray of slope 0, its image w continuously traces out the curve (11.46) which has no multiple points. It follows that the function $w = z + \ln z$ is analytic and one-to-one (i.e., univalent) in the upper half-plane $\operatorname{Im} z > 0$, which it maps conformally onto the domain D obtained from the upper half-plane by excluding the half-line (11.45).

Example. If we set $z = e^t$, the function $w = z + \ln z$ becomes

$$w = e^t + t, \tag{11.47}$$

which is easily seen to be a conformal mapping of the strip $0 < \operatorname{Im} t < \pi$ onto the domain D. Since the function (11.47) takes complex conjugate values at points symmetric with respect to the real axis, it maps the strip $-\pi < \operatorname{Im} t < 0$ onto the domain D^* which is the reflection of D in the real axis. Moreover, (11.47) is a one-to-one mapping of the real axis onto itself, and hence it is a one-to-one mapping of the strip $-\pi < \operatorname{Im} t < \pi$ onto the domain obtained from the w-plane by excluding the two half-lines

$$-\infty < \operatorname{Re} w \leqslant -1, \qquad \operatorname{Im} w = \pi,$$

and

$$-\infty < \operatorname{Re} w \leqslant -1, \qquad \operatorname{Im} w = -\pi.$$

PROBLEMS

11.1. On which of the (unshaded) domains shown in Figure 11.8 can single-valued branches of the function $w = \sqrt{z}$ be defined?

11.2. Verify in detail Remark 3, p. 219.

11.3. It is asserted in the remark on pp. 221–222 that there is no loss of generality in assuming that the two Jordan curves L and L' joining the points z_0 and z_1 have only the points z_0 and z_1 in common. Why is this true?

Hint. L and L' can be replaced by polygonal curves Λ and Λ' joining z_0 to z_1 such that we arrive at z_1 along Λ with the same value of $f(z)$ as if we had gone along L, and similarly for Λ' and L'. Except possibly for segments traversed twice in opposite directions, a closed polygonal curve is the union of a finite number of closed Jordan curves (cf. proof of Theorem 4.12).

11.4. In Example 3, p. 223 find the domains into which the appropriate single-valued branches of the function (11.12) map the domains G', G_U and G_L.

Hint. Use the results of Sec. 51.

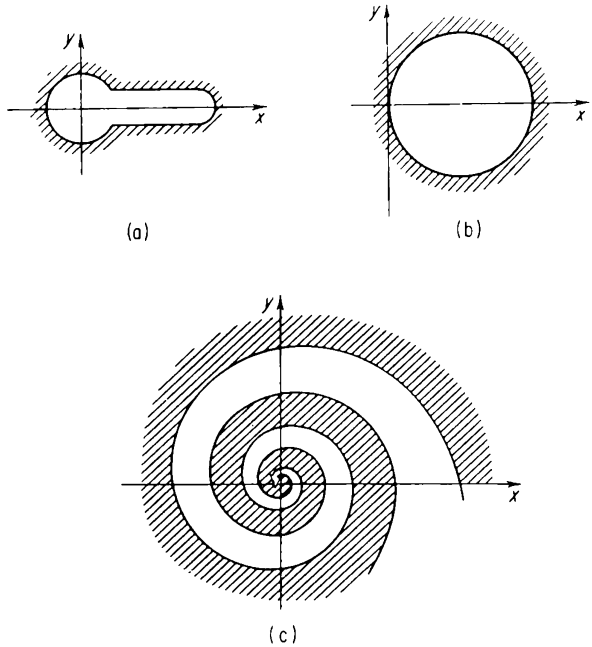

FIGURE 11.8

11.5. Extend the results of Sec. 55 to the more general function

$$w = f(z) = \sqrt[n]{R(z)},$$

where $R(z)$ is an arbitrary rational function.

Hint. To find the branch points of $f(z)$, one must now consider the poles as well as the zeros of $R(z)$.

11.6. Suppose we make a complete circuit in the positive direction around the circle $|z| = 2$, starting and ending at the point $z = 2$. Assuming that the initial value of $\operatorname{Arg} f(z)$ is zero and that $f(z)$ varies continuously with the point z, find the final value of $\operatorname{Arg} f(z)$ if

a) $f(z) = \sqrt{z-1}$; b) $f(z) = \sqrt[3]{z-1}$; c) $f(z) = \sqrt{z^2-1}$;

d) $f(z) = \sqrt{z^2+2z-3}$; e) $f(z) = \sqrt{\dfrac{z-1}{z+1}}$.

Ans. b) $2\pi/3$; d) π.

11.7. Let G be the z-plane cut along the segments $-\infty < x \leqslant -a$ and $a \leqslant x < \infty$ of the real axis.[13] Find a conformal mapping of G onto the half-plane $\operatorname{Im} w > 0$.

---

[13] By the z-plane *cut along* the line segments $\Delta_1, \ldots, \Delta_n$ (finite or infinite), we mean the domain with boundary $\Delta_1 \cup \cdots \cup \Delta_n$ (cf. Problem 10.28).

11.8. Calculate the following logarithms:

a) $\operatorname{Ln} 4$; b) $\operatorname{Ln}(-1)$; c) $\operatorname{Ln} i$; d) $\operatorname{Ln}\frac{1 \pm i}{\sqrt{2}}$; e) $\operatorname{Ln}(2 - 3i)$.

11.9. Suppose we make a complete circuit in the positive direction around the circle $|z| = 2$, starting and ending at the point $z = 2$. Assuming that the value of $\operatorname{Im} f(z)$ is zero at $z = 2$ and that $f(z)$ varies continuously with the point z, find the final value of

a) $f(z) = 2 \operatorname{Ln} z$; b) $f(z) = \operatorname{Ln}\frac{1}{z}$; c) $f(z) = \operatorname{Ln} z - \operatorname{Ln}(z + 1)$;

d) $f(z) = \operatorname{Ln} z + \operatorname{Ln}(z + 1)$.

11.10. Let G be the same domain as in Problem 11.7. Find a conformal mapping of G onto the strip $0 < \operatorname{Im} w < h$, where the left-hand cut goes into the lower edge of the strip and the right-hand cut goes into the upper edge of the strip.

Ans. $w = \frac{h}{\pi} \ln(z + \sqrt{z^2 - a^2}) + hi + c$, where c is an arbitrary constant.

11.11. Find a conformal mapping of the circular lune shown in Figure 11.9 onto the strip $0 < \operatorname{Im} w < h$.

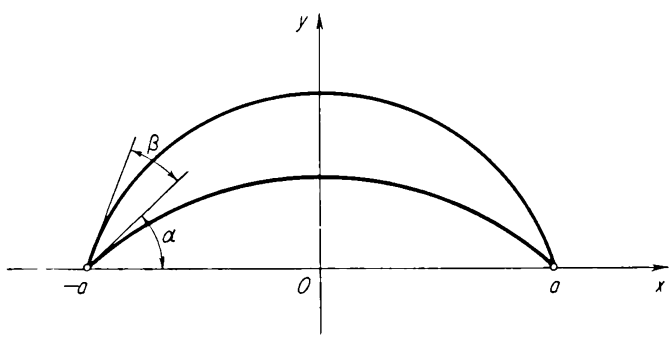

FIGURE 11.9

Ans. $w = \frac{h}{\beta}\left[\ln\frac{z + a}{z - a} - (\pi + \alpha)i\right]$.

11.12. Find all values of the following expressions:

a) $(-2)^{\sqrt{2}}$; b) 2^i; c) 1^{-i}; d) $\left(\frac{1 - i}{\sqrt{2}}\right)^{1+i}$; e) $(3 - 4i)^{1+i}$.

11.13. Calculate $\operatorname{Im} \operatorname{Log}_{1+i}(1 - i)$, choosing as $\operatorname{Ln}(1 + i)$ the principal value of the logarithm.

11.14. Let δ be the segment of the real axis joining the points -1 and $+1$, let I_+ be the positive imaginary axis, and let I_- be the negative imaginary axis. Let D_+ and D_- be the domains with boundaries $\delta \cup I_+$ and $\delta \cup I_-$,

respectively. Define single-valued branches of Arc cos z on each of the domains D_+ and D_-, and find the images of D_+ and D_- corresponding to these branches.

11.15. The *inverse trigonometric functions* Arc sin z, Arc tan z, Arc cot z and the *inverse hyperbolic functions* Arc coth z, Arc sinh z, Arc tanh z and Arc cosh z are defined as the inverses of the corresponding trigonometric functions, just as $w = \text{Arc cos } z$ is defined as the inverse of $z = \cos w$. Verify the following relations:

a) $\text{Arc sin } z = -i \text{ Ln } i(z + \sqrt{z^2 - 1})$;

b) $\text{Arc tan } z = \frac{i}{2} \text{Ln} \frac{i + z}{i - z} = \frac{1}{2i} \text{Ln} \frac{1 + iz}{1 - iz}$; c) $\text{Arc cot } z = \frac{i}{2} \text{Ln} \frac{z - i}{z + i}$;

d) $\text{Arc cosh } z = \text{Ln } (z + \sqrt{z^2 - 1})$; e) $\text{Arc sinh } z = \text{Ln } (z + \sqrt{z^2 + 1})$;

f) $\text{Arc tanh } z = \frac{1}{2} \text{Ln} \frac{1 + z}{1 - z}$; g) $\text{Arc coth } z = \frac{1}{2} \text{Ln} \frac{z + 1}{z - 1}$.

11.16. Prove that the function Arc tan z has only two branch points $\pm i$.

11.17. Let Δ be the finite segment of the imaginary axis joining $-i$ to $+i$, and let Δ' be the infinite segment of the imaginary axis joining $-i$ to $+i$ and passing through the point at infinity. Let D and D' be the domains with boundaries Δ and Δ', respectively. Define single-valued branches of Arc tan z on each of the domains D and D', and find the images of D and D' corresponding to these branches.

Ans. D is mapped onto the strips $k\pi < u < (k + 1)\pi$, and D' is mapped onto the strips $(k - \frac{1}{2})\pi < u < (k + \frac{1}{2})\pi$, where $k = 0, \pm 1, \pm 2, \ldots$

11.18. For which z are all values of the functions Arc cos z, Arc sin z and Arc tan z real? For which z are all values of the function Arc sinh z purely imaginary?

11.19. Find all values of the following expressions:

a) Arc sin $\frac{1}{2}$; b) Arc cos $\frac{1}{2}$; c) Arc cos 2; d) Arc sin i;

e) Arc tan $(1 + 2i)$; f) Arc cosh $2i$; g) Arc tanh $(1 - i)$.

Ans. b) $2k\pi \pm \frac{1}{3}\pi$; d) $2k\pi - i \ln (\sqrt{2} - 1), (2k + 1)\pi - i \ln (\sqrt{2} + 1)$;
f) $\ln (\sqrt{5} \pm 2) + (2k \pm \frac{1}{2})\pi i$.

11.20. Prove that the function

$$f(z) = \frac{1}{(1 - iz)^2}, \quad f(-i) = \infty, \quad f(\infty) = 0$$

is univalent in the upper half-plane Im $z > 0$, but not in the whole complex plane.

11.21. Prove that the function

$$f(z) = \frac{1}{(1 - iz)^3}, \qquad f(\infty) = 0$$

is multivalent in the upper half-plane Im $z > 0$.

Hint. $f(2\sqrt{3} + i) = f(-2\sqrt{3} + i)$.

Part 3

INTEGRATION POWER SERIES

12

RECTIFIABLE CURVES. COMPLEX INTEGRALS

60. Some Basic Definitions

Let L be a continuous curve with equation

$$z = \lambda(t) = \mu(t) + i\nu(t) \qquad (a \leqslant t \leqslant b), \tag{12.1}$$

and suppose we divide the interval $[a, b]$ into n subintervals

$$[t_{k-1}, t_k] \qquad (k = 1, \ldots, n) \tag{12.2}$$

by introducing $n - 1$ intermediate points $t_1, \ldots, t_{n-1}$ satisfying the inequalities

$$a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b.$$

The set

$$\mathscr{P} = \{t_0, t_1, \ldots, t_n\} \tag{12.3}$$

is called a *partition* of the interval $[a, b]$, and the largest of the numbers

$$t_1 - t_0, t_2 - t_1, \ldots, t_n - t_{n-1},$$

i.e., the maximum length of the subintervals (12.2), is called the *norm* of the partition $\mathscr{P}$, denoted by $|\mathscr{P}|$.

Corresponding to the partition (12.3), the curve L is divided into n arcs $\sigma_k = \overset{\frown}{z_{k-1}z_k}$ $(k = 1, \ldots, n)$, where

$$z_k = \lambda(t_k) \qquad (k = 0, 1, \ldots, n),$$

and the final point of each arc (except the last one) coincides with the initial point of the next arc. Joining each of the points $z_0, z_1, \ldots, z_n$ to the next by

a straight line segment (as illustrated by Figure 12.1 for the case $n = 5$), we obtain a polygonal curve Λ *inscribed* in L. The segments of Λ are the chords joining the end points of the arcs σ_k, and the length of Λ is obviously

$$\sum_{k=1}^{n} |z_k - z_{k-1}|. \tag{12.4}$$

The curve L is said to be *rectifiable* if

$$\sup_{\mathscr{P}} \sum_{k=1}^{n} |z_k - z_{k-1}| = l < \infty,$$

where the least upper bound is taken over all possible partitions (12.3). The nonnegative real number l is called the *length* of the curve L. However, if the sums (12.4) become arbitrarily large for suitably chosen partitions, the curve is said to be *nonrectifiable*. In this case, L is considered to have no length at all (or, if preferred, infinite length).

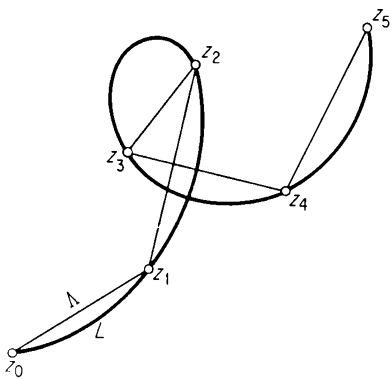

FIGURE 12.1

Remark 1. Let L be a continuous rectifiable curve, and let $\Lambda(\mathscr{P})$ be the polygonal curve inscribed in L corresponding to the partition $\mathscr{P}$. Then it is not hard to show (see Problem 12.3) that the length of $\Lambda(\mathscr{P})$ approaches the length of L as the norm of the partition $\mathscr{P}$ approaches zero.

Remark 2. A continuous curve L is said to be *smooth* if among its various parametric representations, there is at least one representation (12.1) such that $\lambda(t)$ has a continuous nonvanishing derivative $\lambda'(t)$ at every point of the interval $[a, b]$.[1] The geometric meaning of smoothness is clear from Sec. 31, where it was proved that if $\lambda'(t_0) \neq 0$, the curve L has a tangent of inclination $\operatorname{Arg} \lambda'(t_0)$ at the point $z_0 = \lambda(t_0)$. Therefore a smooth curve has a tangent at every point. In fact, as t increases from a to b, z continuously traces out the curve L, and at the same time $\operatorname{Arg} \lambda'(t)$ varies continuously, since $\lambda'(t)$ varies continuously without vanishing.[2] In other words, the inclination of the tangent to a smooth curve varies continuously as the point of tangency moves along the curve. It is not hard to show (see Problem 12.5) that every

[1] At $t = a$, $\lambda'(t)$ denotes the right-hand derivative, and is continuous from the right; at $t = b$, $\lambda'(t)$ denotes the left-hand derivative, and is continuous from the left.

[2] $\operatorname{Arg} \lambda'(t) = \operatorname{Im} \operatorname{Ln} \lambda'(t)$, and since $\lambda'(t)$ varies continuously without vanishing, $\operatorname{Ln} \lambda'(t)$ and hence $\operatorname{Im} \operatorname{Ln} \lambda'(t)$ vary continuously.

smooth curve L is rectifiable, with length l given by the familiar formula

$$l = \int_a^b \sqrt{[\mu'(t)]^2 + [\nu'(t)]^2}\, dt. \tag{12.5}$$

We can write (12.5) in the more concise form

$$l = \int_a^b |\lambda'(t)|\, dt, \tag{12.6}$$

since

$$\lambda'(t) = \mu'(t) + i\nu'(t), \qquad |\lambda'(t)|^2 = [\mu'(t)]^2 + [\nu'(t)]^2.$$

Remark 3. A continuous curve L is said to be *piecewise smooth* if among its various representations, there is at least one representation (12.1) such that the interval $[a, b]$ can be divided into a finite number of subintervals

$$a = a_0 \leqslant t \leqslant a_1, \qquad a_1 \leqslant t \leqslant a_2, \ldots, a_{m-1} \leqslant t \leqslant a_m,$$

on each of which $\lambda(t)$ has a continuous nonvanishing derivative $\lambda'(t)$. In other words, if L is piecewise smooth, then

$$L = \overset{\frown}{Z_0Z_1} + \overset{\frown}{Z_1Z_2} + \cdots + \overset{\frown}{Z_{m-1}Z_m},$$

where $Z_k = \lambda(a_k)$, $k = 0, 1, \ldots, m$, and each of the arcs $\overset{\frown}{Z_{k-1}Z_k}$ is a smooth curve (cf. footnote 12, p. 156). It follows from this definition that although a piecewise smooth curve L may not have a tangent at the points Z_k, it does have a left-hand tangent and a right-hand tangent at each such point (see Problem 8.1), i.e., L has a "corner" at any point where it does not have a tangent. It is an immediate consequence of Problem 12.1 that every piecewise smooth curve L is rectifiable, and that the length of L is still given by formula (12.6).

Of course, piecewise smooth curves are far from being the most general rectifiable curves (cf. Problem 12.6). However, in the treatment that follows, the reader with a limited knowledge of the theory of rectifiable curves can always think in terms of piecewise smooth curves instead of rectifiable curves.

Now let L be an arbitrary rectifiable curve, with equation (12.1), and let

$$P(x, y), \qquad Q(x, y)$$

be two real functions, defined and continuous at every point $z = x + iy$ in L. Given an arbitrary partition $\mathscr{P} = \{t_0, t_1, \ldots, t_n\}$ of the interval $[a, b]$, let σ_k denote the arc $\overset{\frown}{z_{k-1}z_k}$, where

$$z_{k-1} = x_{k-1} + iy_{k-1} = \lambda(t_{k-1}), \qquad z_k = x_k + iy_k = \lambda(t_k),$$

and let

$$\zeta_k = \xi_k + i\eta_k = \lambda(\tau_k) \qquad (t_{k-1} \leqslant \tau_k \leqslant t_k)$$

be an arbitrary point of σ_k ($k = 1, \ldots, n$). Then it can be shown (see Problem 12.11) that the sum

$$\sum_{k=1}^{n} [P(\xi_k, \eta_k)(x_k - x_{k-1}) + Q(\xi_k, \eta_k)(y_k - y_{k-1})] \tag{12.7}$$

approaches a unique limit, denoted by

$$\int_L \{P(x, y)\,dx + Q(x, y)\,dy\}, \tag{12.8}$$

as the norm of the partition $\mathscr{P}$ approaches zero. Expressions of the form (12.8) are known as *line integrals (along the curve L)*. In the next section, we shall use this fact to introduce the concept of the integral of a complex function along a rectifiable curve.

Remark. If L is a piecewise smooth curve with equation (12.1), we can write (12.8) as the following definite integral of a function of the parameter t:

$$\int_a^b \{P[\mu(t), \nu(t)]\mu'(t) + Q[\mu(t), \nu(t)]\nu'(t)\}\,dt.$$

61. Integrals of Complex Functions

Let L be a rectifiable curve with equation (12.1), and let

$$f(z) = u(x, y) + iv(x, y)$$

be a complex function, defined and continuous on L. Given an arbitrary partition $\mathscr{P} = \{t_0, t_1, \ldots, t_n\}$ of the interval $[a, b]$, consider the sum

$$S = \sum_{k=1}^{n} f(\zeta_k)(z_k - z_{k-1}),$$

where

$$\zeta_k = \xi_k + i\eta_k, \quad z_k = \lambda(t_k), \quad \sigma_k = \overset{\frown}{z_{k-1}z_k} \qquad (k = 1, \ldots, n)$$

have the same meaning as in Sec. 60. Introducing the additional notation

$$u_k = u(\xi_k, \eta_k), \quad v_k = v(\xi_k, \eta_k), \quad \Delta x_k = x_k - x_{k-1}, \quad \Delta y_k = y_k - y_{k-1},$$

we have

$$f(\zeta_k) = u_k + iv_k, \qquad z_k - z_{k-1} = \Delta x_k + i\,\Delta y_k,$$

and hence

$$\begin{aligned} S &= \sum_{k=1}^{n} f(\zeta_k)(z_k - z_{k-1}) = \sum_{k=1}^{n} (u_k + iv_k)(\Delta x_k + i\,\Delta y_k) \\ &= \sum_{k=1}^{n} (u_k\,\Delta x_k - v_k\,\Delta y_k) + i\sum_{k=1}^{n} (v_k\,\Delta x_k + u_k\,\Delta y_k). \end{aligned} \tag{12.9}$$

Thus we see that the real and imaginary parts of S are both sums of the form (12.7), corresponding to the partition $\mathscr{P}$, where in the first case the role of the real functions P and Q is played by the functions u and $-v$, and in the second

case by the functions v and u. By the *(contour) integral of* $f(z)$ *along the curve (or contour)* L,[3] denoted by

$$\int_L f(z)\,dz,$$

we mean the quantity

$$\lim_{|\mathscr{P}|\to 0} \sum_{k=1}^{n} f(\zeta_k)(z_k - z_{k-1}), \tag{12.10}$$

i.e., the limit of the sum S as the norm of the partition $\mathscr{P}$ approaches zero. Since the real functions u and v are continuous on L [being the real and imaginary parts of the continuous function $f(z)$], and since L is rectifiable, it follows from (12.9) and the definition of the corresponding line integrals that

$$\int_L f(z)\,dz = \int_L \{u(x, y)\,dx - v(x, y)\,dy\} + i\int_L \{v(x, y)\,dx + u(x, y)\,dy\}. \tag{12.11}$$

Thus we see that evaluating the integral of a complex function reduces to evaluating two line integrals of real functions.

Example 1. If L is the segment $\delta: \alpha \leqslant x \leqslant \beta$ of the real axis, and if $f(z)$ reduces to a real function of the real variable x, then

$$\int_L f(z)\,dz = \int_\delta f(x)\,dx = \lim_{|\mathscr{P}|\to 0} \sum_{k=1}^{n} f(\zeta_k)(x_k - x_{k-1}).$$

Therefore, the definite integral of a real function of a real variable is a special case of the integral of a function of a complex variable along a curve.

Example 2. Again let L be the segment δ, but this time let $f(z) = u(x) + iv(x)$ be a *complex* function of the real variable x. Then we have

$$\begin{aligned}\int_L f(z)\,dz &= \lim_{|\mathscr{P}|\to 0} \sum_{k=1}^{n} [u(\xi_k) + iv(\xi_k)](x_k - x_{k-1}) \\ &= \lim_{|\mathscr{P}|\to 0} \sum_{k=1}^{n} u(\xi_k)(x_k - x_{k-1}) + i \lim_{|\mathscr{P}|\to 0} \sum_{k=1}^{n} v(\xi_k)(x_k - x_{k-1}) \\ &= \int_\alpha^\beta u(x)\,dx + i\int_\alpha^\beta v(x)\,dx. \end{aligned} \tag{12.12}$$

Example 3. If L is a piecewise smooth curve, with equation (12.1), we can replace each of the line integrals in the right-hand side of (12.11) by a suitable definite integral of a complex function of the real parameter t. In fact,

$$\begin{aligned}\int_L f(z)\,dz = \int_a^b \{u[\mu(t), \nu(t)]\mu'(t) - v[\mu(t), \nu(t)]\nu'(t)\}\,dt \\ + i\int_a^b \{v[\mu(t), \nu(t)]\mu'(t) + u[\mu(t), \nu(t)]\nu'(t)\}\,dt,\end{aligned}$$

[3] As in the case of a line integral, L is often called the *path of integration.*

and according to Example 1, the right-hand side is the integral of the function

$$\begin{aligned}\varphi(t) &= \{u[\mu(t), \nu(t)]\mu'(t) - v[\mu(t), \nu(t)]\nu'(t)\} \\ &\quad + i\{v[\mu(t), \nu(t)]\mu'(t) + u[\mu(t), \nu(t)]\nu'(t)\} \\ &= \{u[\mu(t), \nu(t)] + iv[\mu(t), \nu(t)]\}\{\mu'(t) + i\nu'(t)\}\end{aligned}$$

along the segment Δ: $a \leqslant t \leqslant b$ of the real axis. Moreover

$$\varphi(t) = f[\lambda(t)]\lambda'(t),$$

in terms of the function $\lambda(t) = \mu(t) + i\nu(t)$, and hence

$$\int_L f(z)\,dz = \int_a^b \varphi(t)\,dt = \int_a^b f[\lambda(t)]\lambda'(t)\,dt. \tag{12.13}$$

The advantage of the expression on the right is that after separating real and imaginary parts of the integrand, the complex integral reduces to the sum of two definite integrals of real functions of the parameter t.

62. Properties of Complex Integrals

In this section, we derive some elementary properties of integrals of complex functions. The first three properties can be verified immediately either by using formula (12.11) to go over to line integrals, or by using the definition of a complex integral as the limit of the appropriate sum [see (12.10)]. In each case, we assume that L is a rectifiable curve, and that $f(z)$ is defined and continuous on L.

Property 1.

$$\int_L f(z)\,dz = -\int_{-L} f(z)\,dz,$$

where $-L$ denotes the curve L traversed in the opposite direction.

Property 2. Suppose that

$$L = L_1 + L_2 + \cdots + L_m,$$

where the final point of L_k coincides with the initial point of L_{k+1} ($k = 1, 2, \ldots, m - 1$). Then we have

$$\int_L f(z)\,dz = \int_{L_1} f(z)\,dz + \int_{L_2} f(z)\,dz + \cdots + \int_{L_m} f(z)\,dz.$$

Property 3.

$$\int_L [c_1 f_1(z) + \cdots + c_m f_m(z)]\,dz = c_1 \int_L f_1(z)\,dz + \cdots + c_m \int_L f_m(z)\,dz,$$

where $f_1(z), \ldots, f_m(z)$ are arbitrary complex functions which are defined and continuous on L, and $c_1, \ldots, c_m$ are arbitrary complex constants.

Property 4. Suppose $f(z)$ satisfies the inequality

$$|f(z)| \leqslant M$$

at all points of the curve L, e.g., suppose M is the quantity

$$\max_{z \in L} |f(z)|.$$

Then we have the estimate

$$\left| \int_L f(z)\, dz \right| \leqslant Ml, \tag{12.14}$$

where l is the length of the curve L. To prove (12.14), we need only take the limit of the inequality

$$\left| \sum_{k=1}^{n} f(\zeta_k)(z_k - z_{k-1}) \right| \leqslant \sum_{k=1}^{n} |f(\zeta_k)|\, |z_k - z_{k-1}| \leqslant M \sum_{k=1}^{n} |z_k - z_{k-1}| \leqslant Ml$$

as the norm of the partition $\mathscr{P}$ goes to zero.

Remark. All the properties of integrals of complex functions given so far are the exact analogues of the corresponding properties of integrals of real functions. We now indicate an important property of real integrals which cannot be generalized to the case of complex integrals. We refer to the mean value theorem (more exactly, the first law of the mean for integrals), which states that

$$\int_a^b f(t)\, dt = f(\tau) \int_a^b dt = (b - a) f(\tau) \qquad (a < \tau < b), \tag{12.15}$$

where $f(t)$ is a continuous real function defined on the interval $[a, b]$. In particular, (12.15) implies that the integral

$$\int_a^b f(t)\, dt$$

cannot vanish if the function $f(t)$ does not vanish anywhere in the open interval (a, b). However, this last conclusion does not apply to integrals of complex functions, even if we consider only integration along a segment of the real axis. For example, consider the integral

$$\int_L e^{2\pi i x}\, dx$$

taken along the segment $L: 0 \leqslant x \leqslant 1$. Using (12.12), we find that

$$\int_L e^{2\pi i x}\, dx = \int_0^1 \cos 2\pi x\, dx + i \int_0^1 \sin 2\pi x\, dx = 0.$$

But $e^{2\pi i x}$ does not vanish at any point of L. Therefore the conclusion drawn from the mean value theorem does not apply to integrals of complex functions, and hence neither does the mean value theorem itself.

We now give a few examples illustrating the evaluation of the integral

$$\int_L f(z)\,dz, \tag{12.16}$$

where $f(z)$ is a complex function.

Example 1. If $f(z) \equiv 1$, (12.16) becomes

$$\int_L dz = \lim_{|\mathscr{P}|\to 0} \sum_{k=1}^{n} (z_k - z_{k-1}) = \lim_{|\mathscr{P}|\to 0} (z_n - z_0) = Z - z_0,$$

where z_0 is the initial point and Z the final point of the curve L. In particular, if L is closed, $Z = z_0$ and

$$\int_L dz = 0.$$

Example 2. If $f(z) = z$, we first choose ζ_k to be the initial point z_{k-1} of the arc $\sigma_k = \overset{\frown}{z_{k-1}z_k}$, and then the final point z_k of σ_k. This gives

$$\begin{aligned}\int_L z\,dz &= \lim_{|\mathscr{P}|\to 0} \sum_{k=1}^{n} z_{k-1}(z_k - z_{k-1}) = \lim_{|\mathscr{P}|\to 0} \sum_{k=1}^{n} z_k(z_k - z_{k-1}) \\ &= \frac{1}{2} \lim_{|\mathscr{P}|\to 0} \sum_{k=1}^{n} (z_k + z_{k+1})(z_k - z_{k-1}) = \frac{1}{2} \lim_{|\mathscr{P}|\to 0} \sum_{k=1}^{n} (z_k^2 - z_{k-1}^2) \\ &= \frac{1}{2} \lim_{|\mathscr{P}|\to 0} (z_n^2 - z_0^2) = \frac{1}{2}(Z^2 - z_0^2).\end{aligned}$$

In particular, if L is a closed curve, so that $Z = z_0$, we again find that (12.16) vanishes.

Example 3. Suppose $f(z) = 1/(z - a)$, and let L be the curve with equation

$$z = a + re^{it} \qquad (0 \leqslant t \leqslant 2\pi),$$

i.e., the circle of radius r with center at the point a, traversed once in the positive direction. We shall evaluate the integral

$$\int_L f(z)\,dz = \int_L \frac{dz}{z - a}$$

in two different ways. One approach consists in giving the sum

$$\sum_{k=1}^{n} f(\zeta_k)(z_k - z_{k-1})$$

a particularly simple form. We first divide L into n equal arcs $\sigma_k = \overset{\frown}{z_{k-1}z_k}$, where

$$z_k = a + re^{2k\pi i/n} \qquad (k = 0, 1, \ldots, n)$$

($z_n = z_0$, since L is closed), and we then choose ζ_k to be the midpoint of the arc σ_k:

$$\zeta_k = a + re^{(2k-1)\pi i/n} \qquad (k = 1, \ldots, n).$$

As a result,

$$f(\zeta_k) = \frac{1}{\zeta_k - a} = \frac{1}{r} e^{-(2k-1)\pi i/n},$$

and

$$\sum_{k=1}^{n} f(\zeta_k)(z_k - z_{k-1}) = \sum_{k=1}^{n} e^{-(2k-1)\pi i/n}(e^{2k\pi i/n} - e^{2(k-1)\pi i/n})$$
$$= \sum_{k=1}^{n} (e^{\pi i/n} - e^{-\pi i/n}) = 2i \sum_{k=1}^{n} \sin\frac{\pi}{n} = 2in \sin\frac{\pi}{n}.$$

It follows that

$$\int_L \frac{dz}{z-a} = \lim_{n\to\infty} 2in \sin\frac{\pi}{n} = 2\pi i \lim_{n\to\infty} \frac{\sin(\pi/n)}{\pi/n} = 2\pi i.$$

Another way of obtaining the same result is to use formula (12.13). In fact, since

$$z = \lambda(t) = a + re^{it}, \qquad \lambda'(t) = ire^{it},$$

we have

$$\int_L \frac{dz}{z-a} = \int_0^{2\pi} \frac{ire^{it}\,dt}{re^{it}} = i\int_0^{2\pi} dt = 2\pi i.$$

PROBLEMS

12.1. Given a continuous curve L with end points A and B, let C be any other point of L. Prove that the curve $L = \overset{\frown}{AB}$ is rectifiable if and only if the subcurves $\overset{\frown}{AC}$ and $\overset{\frown}{CB}$ are rectifiable. Assuming that these curves are rectifiable, prove that the length of $\overset{\frown}{AB}$ is the sum of the lengths of $\overset{\frown}{AC}$ and $\overset{\frown}{CB}$.

12.2. Given a continuous curve L with equation $z = \lambda(t)$, $a \leqslant t \leqslant b$, let $\mathscr{P} = \{t_0, t_1, \ldots, t_n\}$ and $\tilde{\mathscr{P}} = \{\tilde{t}_0, \tilde{t}_1, \ldots, \tilde{t}_{m+1}\}$, $m \geqslant 1$, be two partitions of the interval $[a, b]$. Let $\mathscr{P}^*$ be the partition consisting of the points of $\mathscr{P}$ and the points of $\tilde{\mathscr{P}}$ not already included in $\mathscr{P}$. Prove that

$$l(\mathscr{P}) \leqslant l(\mathscr{P}^*), \qquad l(\tilde{\mathscr{P}}) \leqslant l(\mathscr{P}^*)$$

and

$$l(\mathscr{P}^*) \leqslant l(\mathscr{P}) + 2m \max_{|t-t'| \leqslant |\mathscr{P}|} |\lambda(t) - \lambda(t')|,$$

where $l(\mathscr{P})$, $l(\tilde{\mathscr{P}})$ and $l(\mathscr{P}^*)$ are the lengths of the polygonal curves inscribed in L corresponding to the partitions $\mathscr{P}$, $\tilde{\mathscr{P}}$ and $\mathscr{P}^*$, respectively.

12.3. Prove the assertion made in Remark 1, p. 246.

Hint. Use the preceding problem.

12.4. Give an example of a curve $z = \lambda(t)$, $a \leqslant t \leqslant b$, with continuously differentiable $\lambda(t)$, which fails to have a tangent at some point of $[a, b]$.

Hint. Let $\lambda(t) = t^3 + i|t^3|$ on the interval $[-1, 1]$.

Comment. In Remark 2, p. 246, this difficulty is prevented by the assumption that $\lambda'(t)$ is nonvanishing on $[a, b]$.

12.5. Let L be a smooth curve with equation $z = \lambda(t) = \mu(t) + i\nu(t), a \leqslant t \leqslant b$. Prove that L is rectifiable, with length given by formula (12.5).

Hint. Use the mean value theorem to write

$$\begin{aligned}\sum_{k=1}^{n} |z_k - z_{k-1}| &= \sum_{k=1}^{n} \sqrt{[\mu(t_k) - \mu(t_{k-1})]^2 + [\nu(t_k) - \nu(t_{k-1})]^2} \\ &= \sum_{k=1}^{n} \sqrt{[\mu'(\tau_k)]^2 + [\nu'(\tilde{\tau}_k)]^2}\, \Delta t_k,\end{aligned} \tag{12.17}$$

where τ_k and $\tilde{\tau}_k$ are two points in the interval $[t_{k-1}, t_k]$, and $\Delta t_k = t_k - t_{k-1}$. Except for the fact that τ_k and $\tilde{\tau}_k$ are in general different, (12.17) is an approximating sum of the Riemann integral (12.5). Use the fact that $\mu'(t)$ and $\nu'(t)$ are continuous to circumvent this difficulty.

Comment. Note that we have not used the fact that $\lambda'(t)$ is nonvanishing on $[a, b]$.

12.6. A real function $f(t)$ defined on a closed interval $[a, b]$ is said to be of *bounded variation* on $[a, b]$ if

$$\sup_{\mathscr{P}} \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| < \infty,$$

where the supremum if taken over all possible partitions $\mathscr{P} = \{t_0, t_1, \ldots, t_n\}$ of the interval $[a, b]$. Prove that the curve with equation $z = \lambda(t) = \mu(t) + i\nu(t)$, $a \leqslant t \leqslant b$, is rectifiable if and only if both $\mu(t)$ and $\nu(t)$ are of bounded variation on $[a, b]$.

12.7. Show that the function

$$f(t) = \begin{cases} t \sin \dfrac{\pi}{2t} & \text{for } t \neq 0, \\ 0 & \text{for } t = 0 \end{cases}$$

is continuous but not of bounded variation on $[0, 1]$. Give an example of a continuous nonrectifiable curve.

12.8. Prove that if $f(t)$ is continuously differentiable on $[a, b]$, then $f(t)$ is of bounded variation on $[a, b]$.

Comment. Together with Problem 12.6, this gives another proof of the fact that a smooth curve is rectifiable (cf. Problem 12.5).

12.9. Prove that $f(t)$ is of bounded variation on $[a, b]$ if and only if

$$f(t) = f_1(t) - f_2(t)$$

on $[a, b]$, where $f_1(t)$ and $f_2(t)$ are both increasing functions on $[a, b]$ (see Problem 4.7).

Hint. First prove that $f(t)$ is of bounded variation on $[a, b]$ if and only if

$f(t)$ has a *majorant* $F(t)$ on $[a, b]$, i.e., a bounded increasing function $F(t)$ such that

$$|f(t'') - f(t')| \leqslant F(t'') - F(t')$$

whenever $a \leqslant t' < t'' \leqslant b$.

12.10. Let $f(t)$ and $g(t)$ be two functions defined on the same interval $[a, b]$, let $\mathscr{P} = \{t_0, t_1, \ldots, t_n\}$ be an arbitrary partition of $[a, b]$, and let τ_k be any point in the interval $[t_{k-1}, t_k]$. Then the limit

$$\lim_{|\mathscr{P}| \to 0} \sum_{k=1}^{n} f(\tau_k)[g(t_k) - g(t_{k-1})],$$

provided it exists, is called the *Stieltjes integral* of $f(t)$ with respect to $g(t)$ from a to b, and is denoted by

$$\int_a^b f(t)\, dg(t).$$

Prove that this limit exists if $f(t)$ is continuous on $[a, b]$ and $g(t)$ is of bounded variation on $[a, b]$.

Hint. First consider the case where $g(t)$ is increasing on $[a, b]$, and repeat step by step the familiar proof that the ordinary Riemann integral

$$\int_a^b f(t)\, dt$$

exists if $f(t)$ is continuous on $[a, b]$. Then use the preceding problem.[4]

12.11. Let L be a continuous rectifiable curve, with equation

$$z = \lambda(t) = \mu(t) + i\nu(t) \qquad (a \leqslant t \leqslant b),$$

and let $P(x, y)$ and $Q(x, y)$ be two real functions which are defined and continuous on L. Prove the existence of the line integral

$$\int_L \{P(x, y)\, dx + Q(x, y)\, dy\},$$

as defined on p. 248.

Hint. Write the line integral as a sum of Stieltjes integrals with respect to $\mu(t)$ and $\nu(t)$, and use Problems 12.6 and 12.10.

12.12. Evaluate the integrals

$$J_1 = \int_L x\, dz, \qquad J_2 = \int_L y\, dz$$

along the following paths:

a) The line segment joining the point $z = 0$ to the point $z = 2 + i$;
b) The semicircle $|z| = 1$, $0 \leqslant \arg z \leqslant \pi$, with initial point $z = i$;
c) The circle $|z - a| = R$.

Ans. b) $J_1 = \frac{1}{2}\pi i$, $J_2 = -\frac{1}{2}\pi i$.

---

[4] The details for increasing $g(t)$ are given in D. V. Widder, *op. cit.*, p. 176.

12.13. Evaluate the integral

$$\int |z|\, dz$$

along

a) The same paths as in the preceding problem;

b) The semicircle $|z| = 1$, $-\pi/2 \leqslant \arg z \leqslant \pi/2$ with initial point $z = -i$.

12.14. Evaluate the integral

$$\int_L |z|\bar{z}\, dz,$$

where L is the closed contour consisting of the line segment $-1 \leqslant x \leqslant 1$, $y = 0$ and the semicircle $|z| = 1$, $y \geqslant 0$.

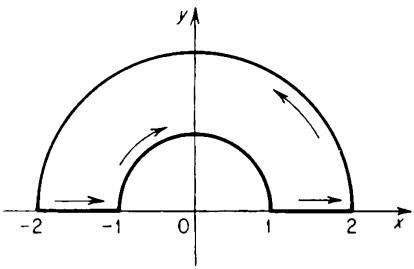

FIGURE 12.2

12.15. Evaluate the integral

$$\int_L \frac{z}{\bar{z}}\, dz,$$

where L is the contour shown in Figure 12.2.

Ans. $\frac{4}{3}$.

12.16. Prove that

$$\lim_{r \to 0} \int_0^{2\pi} f(re^{i\varphi})\, d\varphi = 2\pi f(0),$$

if $f(z)$ is continuous in a neighborhood of the origin $z = 0$.

12.17. Prove that

$$\lim_{r \to 0} \int_{|z-a|=r} \frac{f(z)}{z - a}\, dz = 2\pi i f(a),$$

if $f(z)$ is continuous in a neighborhood of the point $z = a$.

12.18. Prove that if $f(z)$ is continuous in the closed domain $|z| \geqslant R_0$, $0 \leqslant \arg z \leqslant \alpha$ $(0 < \alpha \leqslant 2\pi)$, and if the limit

$$\lim_{r \to \infty} zf(z) = A$$

exists, then

$$\lim_{r \to \infty} \int_{\Gamma_R} f(z)\, dz = iA\alpha,$$

where Γ_R is the arc of the semicircle $|z| = R$ lying in the given domain and traversed in the positive direction.

12.19. Let L be a rectifiable curve with equation $z = \lambda(t)$, $a \leqslant t \leqslant b$, and let $s(t)$ be the length of the arc of L joining $\lambda(a)$ to $\lambda(t)$ [cf. Problem 12.1], where $s(a) = 0$ by definition. Prove that the nonnegative function $s(t)$ is continuous and increasing on $[a, b]$, in fact, *strictly* increasing if L is a Jordan curve.

12.20. Let $f(z)$ be a continuous function on a rectifiable curve L of length l, with equation $z = \lambda(t)$, $a \leqslant t \leqslant b$, and let $s(t)$ be the arc length function

introduced in the preceding problem. By $\int_L f(z)\,ds$ or $\int_L f(z)\,|dz|$, we mean the Stieltjes integral

$$\int_a^b f[\lambda(t)]\,ds(t),$$

whose existence is guaranteed by the fact that $s(t)$ is of bounded variation on $[a, b]$. Prove that

$$\left|\int_L f(z)\,dz\right| \leqslant \int_L |f(z)|\,ds \leqslant Ml,$$

which sharpens the estimate (12.14).

Comment. As in Sec. 60, let $\mathscr{P} = \{t_0, t_1, \ldots, t_n\}$ be a partition of the interval $[a, b]$, let σ_k be the arc of L joining $\lambda(t_{k-1})$ to $\lambda(t_k)$, and let ζ_k be any point of σ_k $(k = 1, \ldots, n)$. Then

$$\int_L f(z)\,ds = \lim_{|\mathscr{P}|\to 0} \sum_{k=1}^{n} f(\zeta_k) l_k,$$

where l_k is the length of σ_k.

12.21. Consider a rectifiable curve L in the Lobachevskian plane (see Sec. 50). Prove that the $\mathscr{L}$-length of L is given by the integral

$$\int_L \frac{ds}{1 - |z|^2}.$$

Hint. In Problem 10.23, let $z_1 = z$, $z_2 = z + \Delta z$.

13

CAUCHY'S INTEGRAL THEOREM

63. A Preliminary Result

This chapter is devoted to the proof of the following theorem, which is one of the most important results in the theory of analytic functions:

THEOREM 13.1 (*Cauchy's integral theorem*). *Let G be a simply connected domain,*[1] *and let $f(z)$ be a single-valued analytic function on G. Then*

$$\int_L f(z)\,dz = 0,$$

where L is any closed rectifiable curve contained in G.

The proof of Theorem 13.1 is rather complicated, and will be given in the next two sections. We begin by proving the following weaker version of Theorem 13.1:

THEOREM 13.2. *Let G be a simply connected domain, and let $f(z)$ be a single-valued continuously differentiable function on G. Then*

$$\int_L f(z)\,dz = 0,$$

where L is any closed piecewise smooth Jordan curve contained in G.

[1] From now on, we adhere to the definition of a simply connected domain given on page 72, i.e., a domain G, bounded or not, is said to be simply connected if whenever G contains a closed Jordan curve γ, G also contains $I(\gamma)$ [the interior of γ].

Proof. According to formula (12.11),

$$\int_L f(z)\,dz = \int_L (u\,dx - v\,dy) + i\int_L (v\,dx + u\,dy), \tag{13.1}$$

where $u = \operatorname{Re} f(z)$, $v = \operatorname{Im} f(z)$. Since $f'(z)$ is continuous on G and

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y},$$

it follows that the first-order partial derivatives of u and v are continuous on $\bar{D}$, where $D = I(L)$ is the interior of an arbitrary closed piecewise smooth Jordan curve $L \subset G$. We now apply Green's theorem[2]

$$\int_L (P\,dx + Q\,dy) = \iint_{\bar{D}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy \tag{13.2}$$

to each of the integrals in the right-hand side of (13.1), where the line integral in (13.2) is taken in the positive (i.e., counterclockwise) direction. The result is

$$\begin{aligned} \int_L (u\,dx - v\,dy) &= \iint_{\bar{D}} \left(-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) dx\,dy, \\ \int_L (v\,dx + u\,dy) &= \iint_{\bar{D}} \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) dx\,dy. \end{aligned} \tag{13.3}$$

But the integrand of each double integral in (13.3) vanishes, because of the Cauchy-Riemann equations (see Theorem 7.2). Therefore both sides of (13.1) vanish, and the proof is complete.

Remark 1. The strengthening of Theorem 13.2 represented by Theorem 13.1 consists in dropping the requirement that $f'(z)$ be continuous and that L be a Jordan curve, and in assuming that L is only rectifiable instead of piecewise smooth.

Remark 2. Actually, it turns out (see Sec. 72) that $f'(z)$ is analytic (and hence automatically continuous) on G, if $f(z)$ is analytic on G. However, Theorem 12.1 will be needed to prove this!

64. The Key Lemma

The following result plays a basic role in the proof of Theorem 13.1:

LEMMA (*Key lemma*). *Let $f(z)$ be a continuous function defined on a domain G, and let L be an arbitrary rectifiable curve contained in G, with equation $z = \lambda(t)$, $a \leqslant t \leqslant b$. Then, given any $\varepsilon > 0$, there exists a*

[2] See e.g., D. V. Widder, *op. cit.*, p. 223.

$\delta = \delta(\varepsilon) > 0$ *such that for every partition* $\mathscr{P} = \{t_0, t_1, \ldots, t_n\}$ *of the interval* $[a, b]$ *with norm less than* δ, *the polygonal curve*

$$\Lambda = \overline{z_0 z_1} + \cdots + \overline{z_{n-1} z_n}, \tag{13.4}$$

where $z_k = \lambda(t_k)$, $k = 0, 1, \ldots, n$ *is contained in* G, *and moreover*

$$\left| \int_L f(z)\, dz - \int_\Lambda f(z)\, dz \right| < \varepsilon. \tag{13.5}$$

Proof. The curve L is a compact set (see Theorem 3.15). Therefore, according to Theorem 4.15, there is another compact set $E \subset G$ and a number $\rho > 0$ such that E contains all the closed neighborhoods $\overline{\mathscr{N}(z; \rho)}$, $z \in L$. Since $f(z)$ is uniformly continuous on E (see Theorem 3.17), given any $\varepsilon > 0$, there exists an $\eta > 0$ such that

$$|f(z) - f(z')| < \frac{\varepsilon}{2l} \tag{13.6}$$

for every pair of points z, z' in E such that $|z - z'| < \eta$, where l is the length of L. Moreover, the function $z = \lambda(t)$ is uniformally continuous on $[a, b]$, and hence there exists a $\delta_1 > 0$ such that

$$|z - z'| = |\lambda(t) - \lambda(t')| < \min(\eta, \rho)$$

if $|t - t'| < \delta_1$.

Now suppose

$$\left| \int_L f(z)\, dz - \sum_{k=1}^{n} f(z_k)(z_k - z_{k-1}) \right| < \frac{\varepsilon}{2} \tag{13.7}$$

if $|\mathscr{P}| < \delta_2$ (such a δ_2 exists by the very definition of the complex integral) and let $\delta = \min(\delta_1, \delta_2)$. Choose a partition $\mathscr{P} = \{t_0, t_1, \ldots, t_n\}$ with norm less than δ, and let Λ be the corresponding polygonal curve (13.4). Clearly $\Lambda \subset G$, since if Δ_k is the segment $\overline{z_{k-1} z_k}$, then $z \in \Delta_k$ implies

$$|z - z_k| \leqslant |z_k - z_{k-1}| \leqslant \min(\eta, \rho) \leqslant \rho,$$

and hence $z \in \overline{\mathscr{N}(z_k; \rho)} \subset E \subset G$. Furthermore, as we now prove,

$$\left| \int_\Lambda f(z)\, dz - \sum_{k=1}^{n} f(z_k)(z_k - z_{k-1}) \right| < \frac{\varepsilon}{2}.$$

In fact,

$$f(z_k)(z_k - z_{k-1}) = f(z_k) \int_{\Delta_k} dz = \int_{\Delta_k} f(z_k)\, dz,$$

so that

$$\left|\int_\Lambda f(z)\,dz - \sum_{k=1}^{n} f(z_k)(z_k - z_{k-1})\right| = \left|\sum_{k=1}^{n}\int_{\Delta_k} f(z)\,dz - \sum_{k=1}^{n}\int_{\Delta_k} f(z_k)\,dz\right|$$
$$= \left|\sum_{k=1}^{n}\int_{\Delta_k} [f(z) - f(z_k)]\,dz\right|$$
$$\leqslant \sum_{k=1}^{n}\left|\int_{\Delta_k} [f(z) - f(z_k)]\,dz\right|. \quad (13.8)$$

But if $z \in \Delta_k$, then, as already noted, both z and z_k belong to E, and moreover

$$|z - z_k| \leqslant |z_k - z_{k-1}| < \min(\eta, \rho) \leqslant \eta,$$

which implies

$$|f(z) - f(z_k)| < \frac{\varepsilon}{2l}$$

[cf. (13.6)]. It follows from the general estimate (12.14) that

$$\left|\int_{\Delta_k} [f(z) - f(z_k)]\,dz\right| \leqslant \frac{\varepsilon}{2l}|z_k - z_{k-1}|,$$

and hence (13.8) becomes

$$\left|\int_\Lambda f(z)\,dz - \sum_{k=1}^{n} f(z_k)(z_k - z_{k-1})\right| \leqslant \sum_{k=1}^{n}\frac{\varepsilon}{2l}|z_k - z_{k-1}| < \frac{\varepsilon}{2} \quad (13.9)$$

(recall the definition of l). Comparing (13.9) and (13.7), we obtain the inequality (13.5), and the proof is complete.

65. Proof of Cauchy's Integral Theorem

We now turn to the proof proper of Theorem 13.1 (Cauchy's integral theorem). First we prove the theorem for closed polygonal curves, and then we use the lemma of the preceding section to prove the theorem for arbitrary closed rectifiable curves. The proof will be divided into six steps, of progressively increasing complexity.

STEP 1 ("*2-gons*"). If L is a straight line segment Δ, traversed twice in opposite directions, then obviously

$$\int_L f(z)\,dz = \int_\Delta f(z)\,dz + \int_{-\Delta} f(z)\,dz = \int_\Delta f(z)\,dz - \int_\Delta f(z)\,dz = 0.$$

Thus, in this simplest case, the differentiability of the function need not be invoked at all.

STEP 2 (*Triangles*). As we shall see in the course of the proof, this is the basic step, and here essential use will be made of the fact that $f(z)$ is

differentiable. Let L be a triangular contour contained in G, and suppose L is traversed once in a definite direction (say counterclockwise). Moreover, let

$$\left|\int_L f(z)\,dz\right| = M,$$

where obviously $M \geqslant 0$. Our goal is to show that $M = 0$. To this end, we draw the line segments joining midpoints of the sides of the triangle, thereby obtaining four congruent subtriangles L^{I}, L^{II}, L^{III} and L^{IV}, as shown in Figure 13.1.[3]

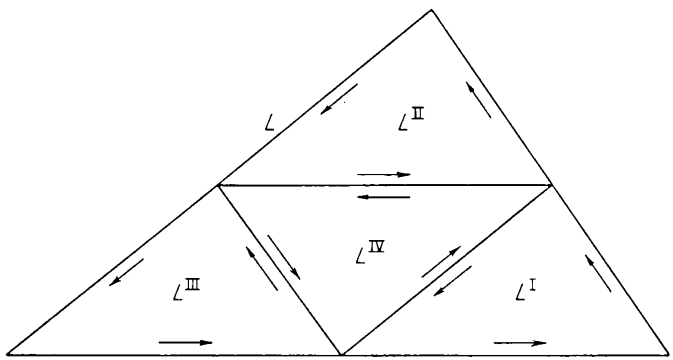

FIGURE 13.1

Forming the sum of the integrals taken along L^{I}, L^{II}, L^{III} and L^{IV} in the directions shown by the arrows in the figure (i.e., in the counterclockwise direction), we obtain

$$\int_{L^{\mathrm{I}}} f(z)\,dz + \int_{L^{\mathrm{II}}} f(z)\,dz + \int_{L^{\mathrm{III}}} f(z)\,dz + \int_{L^{\mathrm{IV}}} f(z)\,dz. \qquad (13.10)$$

Each of these four integrals can be replaced by the sum of three integrals taken along the separate sides of the subtriangles. Of the resulting integrals, six have a sum equal to $\int_L f(z)\,dz$ (i.e., the integrals along the segments whose union is L). The remaining six integrals can be grouped into three pairs, where each pair is taken along the same segment traversed twice in opposite directions. According to Step 1, the sum of each of these pairs of integrals equals zero. Therefore the whole sum (13.10) equals $\int_L f(z)\,dz$, and moreover

[3] For simplicity, we shall use the same symbols L, L^{I}, L_1, etc., (and the same word "triangle") to denote both triangular contours and the closed domains bounded by such contours. Whether we are talking about contours or domains will always be clear from the context. (The term *contour* is used as a synonym for *curve*, mainly when the curve is closed.)

$$M = \left|\int_L f(z)\,dz\right| \leqslant \left|\int_{L^{\mathrm{I}}} f(z)\,dz\right| + \left|\int_{L^{\mathrm{II}}} f(z)\,dz\right| + \left|\int_{L^{\mathrm{III}}} f(z)\,dz\right| + \left|\int_{L^{\mathrm{IV}}} f(z)\,dz\right|, \tag{13.11}$$

since the modulus of a sum does not exceed the sum of the moduli of the summands. It follows from (13.11) that at least one of the terms in the right-hand side must be no less than $M/4$. Denoting the corresponding triangle by L_1 (so that L_1 is one of the triangles L^{I}, L^{II}, L^{III}, L^{IV}), we have

$$\left|\int_{L_1} f(z)\,dz\right| \geqslant \frac{M}{4}.$$

Next we apply the same argument to the triangle L_1 as to the original triangle L. Thus, dividing L_1 into four congruent subtriangles L_1^{I}, L_1^{II}, L_1^{III} and L_1^{IV}, we note that the integral along L_1 equals the sum of four integrals along L_1^{I}, L_1^{II}, L_1^{III} and L_1^{IV}, taken in the same (counterclockwise) direction. It follows that the modulus of one of these integrals is no less than

$$\frac{1}{4} \cdot \frac{M}{4} = \frac{M}{4^2}.$$

If this integral is along L_2 (so that L_2 is one of the triangles L_1^{I}, L_1^{II}, L_1^{III}, L_1^{IV}), we have

$$\left|\int_{L_2} f(z)\,dz\right| \geqslant \frac{M}{4^2}.$$

Continuing this argument indefinitely, we obtain a sequence of triangles $L = L_0, L_1, L_2, \ldots, L_n, \ldots$ such that

1. L_{n+1} is a proper subset of L_n ($n = 0, 1, 2, \ldots$), and in fact L_{n+1} is obtained by drawing the line segments joining the midpoints of two sides of L_n.
2. If $l = l_0$ is the length of $L = L_0$ and l_n the length of L_n, then
$$l_n = \frac{l}{2^n}.$$
3. Each triangle is contained in G, since the curve $L \subset G$ and hence $I(L) \subset G$ (cf. footnote 1, p. 258).
4. The integral of $f(z)$ along L_n satisfies the inequality
$$\left|\int_{L_n} f(z)\,dz\right| \geqslant \frac{M}{4^n} \qquad (n = 1, 2, \ldots). \tag{13.12}$$

It follows from the first and second of the properties just enumerated that there is a unique point ζ belonging to all the triangles, and from the third property that $\zeta \in G$. Therefore, by hypothesis, the function $f(z)$ has

a derivative $f'(\zeta)$ at the point ζ (cf. the statement of Theorem 13.1), i.e., for any $\varepsilon > 0$, we can find a $\delta > 0$ such that

$$|f(z) - f(\zeta) - f'(\zeta)(z - \zeta)| < \varepsilon|z - \zeta| \tag{13.13}$$

whenever $|z - \zeta| < \delta$ (cf. Theorem 7.1). Since ζ belongs to all the "nested" triangles, and since the triangles become arbitrarily small ($l_n \to 0$ as $n \to \infty$), then, starting from some value of $n > N$, the triangles are all contained in the disk $|z - \zeta| < \delta$, so that all their points satisfy the inequality (13.13).

We now integrate the function $f(z) - f(\zeta) - f'(\zeta)(z - \zeta)$ along the closed curve L_n, obtaining

$$\begin{aligned}\int_{L_n} [f(z) - f(\zeta) - f'(\zeta)(z - \zeta)]\, dz &= \int_{L_n} f(z)\, dz - f(\zeta)\int_{L_n} dz - f'(\zeta)\int_{L_n} z\, dz + f'(\zeta)\int_{L_n} dz \\ &= \int_{L_n} f(z)\, dz,\end{aligned}$$

where we have used the fact that the integrals $\int_{L_n} dz$ and $\int_{L_n} z\, dz$ vanish (see Sec. 62, Examples 1 and 2). Therefore, according to (13.13) and (12.13), if $n > N$,

$$\left|\int_{L_n} f(z)\, dz\right| = \left|\int_{L_n} [f(z) - f(\zeta) - f'(\zeta)(z - \zeta)]\, dz\right| < \varepsilon l_n^2 = \varepsilon \frac{l^2}{4^n}, \tag{13.14}$$

where, in writing the inequality, we note that the distance between any two points of the same triangle must be less than the perimeter of the triangle. Comparing (13.12) and (13.14), we find that

$$\frac{M}{4^n} \leqslant \left|\int_{L_n} f(z)\, dz\right| < \frac{\varepsilon l^2}{4^n},$$

which implies

$$M < \varepsilon l^2,$$

or

$$M \leqslant 0$$

if we let $\varepsilon \to 0$. But $M \geqslant 0$, since M is intrinsically nonnegative. Therefore

$$M = \left|\int_L f(z)\, dz\right| = 0,$$

and the proof of Cauchy's integral theorem for triangles is complete.

STEP 3 (*Convex polygons*). Next let $L = A_0A_1 \cdots A_{n-1}A_0$ ($n \geqslant 4$) be the boundary of an arbitrary convex n-gon (see Problem 4.4), traversed in a definite direction.[4] If we divide the polygon into $n - 2$ triangles by drawing

[4] By A_iA_j we mean an arc (or line segment) joining A_i to A_j, where the arc in question is always clear from the context. By $A_{i_1}A_{i_2}A_{i_3} \cdots A_{i_{n-1}}A_{i_n}$ we mean the oriented curve $A_{i_1}A_{i_2} + A_{i_2}A_{i_3} + \cdots + A_{i_{n-1}}A_{i_n}$.

diagonals from the vertex A_0 to the other vertices, the integral of $f(z)$ along L can be written in the form

$$\int_L f(z)\,dz = \int_{A_0A_1A_2A_0} f(z)\,dz + \int_{A_0A_2A_3A_0} f(z)\,dz + \cdots + \int_{A_0A_{n-2}A_{n-1}A_0} f(z)\,dz, \tag{13.15}$$

where each of the triangular contours $A_0A_kA_{k+1}A_0$ $(k = 1, \ldots, n-2)$ is traversed in the same direction (as illustrated by Figure 13.2 for the case $n = 6$). But according to Step 2, each of the integrals in the right-hand side of (13.15) vanishes.[5] This proves Cauchy's theorem for convex polygons.

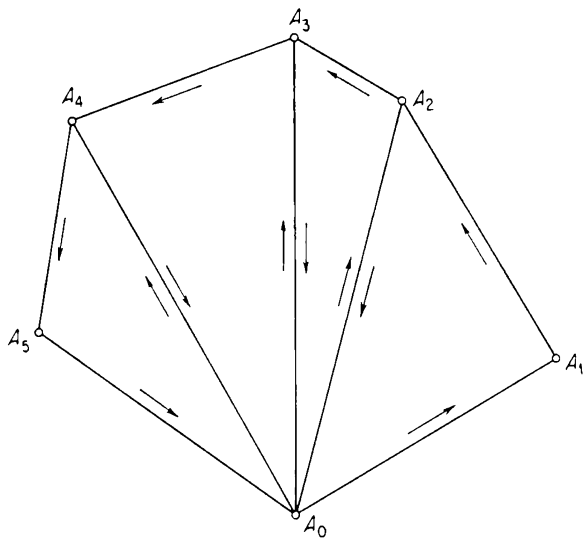

FIGURE 13.2

STEP 4 (*Closed polygonal Jordan curves*). Now let $L = A_0A_1 \cdots A_{n-1}A_0$ $(n \geq 4)$ be any closed polygonal curve (traversed once in a definite direction), which does not intersect itself. Then L is a Jordan curve, and $\overline{I(L)} \subset G$. The case where $\overline{I(L)}$ is convex has already been considered in Step 3. If $\overline{I(L)}$ is not convex, the extensions of some of the sides

$$A_0A_1, A_1A_2, \ldots, A_{n-1}A_0$$

in one direction or the other (possibly both) intersect $I(L)$, as illustrated by Figure 13.3 for the case $n = 6$ (see Problem 13.3). Extending each side $A_{k-1}A_k$ $(k = 1, \ldots, n;\ A_n = A_0)$ which intersects $I(L)$, and terminating the extension of $A_{k-1}A_k$ (in the appropriate direction) at the point where it first

[5] Note that $A_0A_kA_{k+1}A_0 \subset G$, since $L \subset G$ and $I(L) \subset G$.

intersects the boundary L, we decompose the original boundary into a finite number of subpolygons, each of which is convex (why?). Since $\int_L f(z)\,dz$ is obviously the sum of the integrals of $f(z)$ along the boundaries of these convex subpolygons, the fact that $\int_L f(z)\,dz$ vanishes follows at once from Step 3.

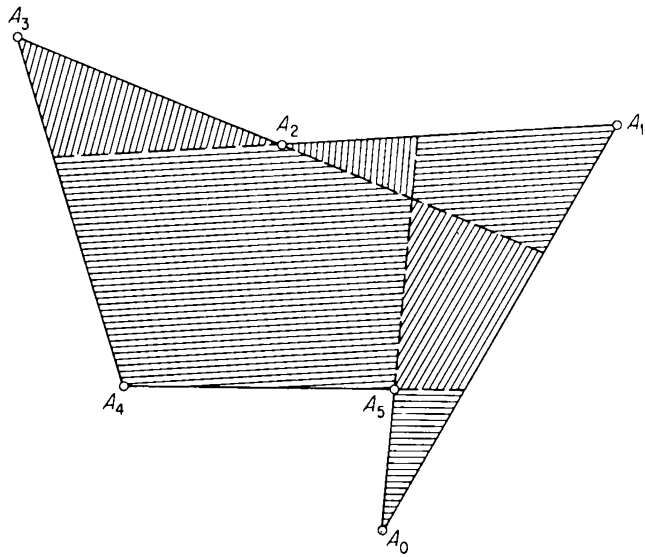

FIGURE 13.3

STEP 5 (*Arbitrary closed polygonal curves*). Next let L be an arbitrary closed polygonal curve. By definition (see p. 61), L consists of a finite number of line segments $\Delta_1, \Delta_2, \ldots, \Delta_n$ with definite directions, such that the final point of each segment Δ_k coincides with the initial point of the next segment Δ_{k+1} ($\Delta_{n+1} = \Delta_1$). Some of these segments may have points in common other than those just indicated, i.e., the curve may intersect itself. In this case, some of the segments Δ_k may be parts of other segments, or may even coincide with other segments. In other words, as we go along the polygonal curve L, some of its segments may be partially or totally traversed several times.

To keep the notation simple, we shall study the polygonal curve shown in Figure 13.4, consisting of eight segments $\Delta_k = A_{k-1}A_k$ ($k = 1, \ldots, 8$; $A_8 = A_0$), where $A_6A_7 \subset A_5A_6$. Starting from A_0, we move along L until a new segment intersects a previously traversed segment for the first time. In our case, A_2A_3 is such a segment, since it intersects A_0A_1 in the point B. Then the closed polygonal curve (here, the triangle BA_1A_2B), obtained by starting from the point B and traversing L until the *first* return to the point

B, is a closed Jordan curve contained in G. Therefore

$$\int_{BA_1A_2B} f(z)\,dz = 0,$$

and we do not change the value of the integral of $f(z)$ along L if we remove BA_1A_2B from L, thereby obtaining a new polygonal curve

$$L' = A_0BA_3A_4A_5A_6A_7A_0,$$

with one vertex less than L.[6]

There is still another possibility which must be considered. Starting from A_0 and moving along the new polygonal curve L', we find that L' does not intersect itself before the point A_6. After A_6 we encounter previously traversed

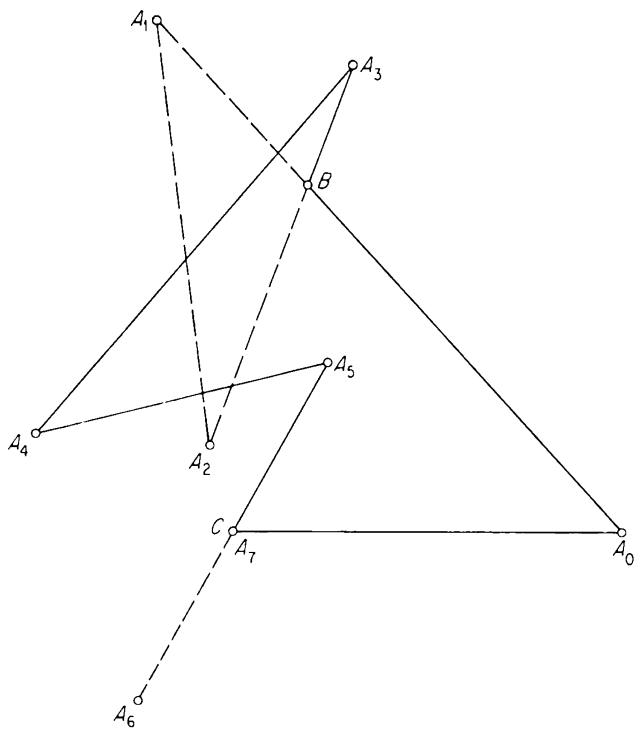

FIGURE 13.4

points, since A_6A_7 is a part of A_5A_6, but none of these points can be called the first, since the point A_6 is a common end point of the two consecutive segments A_5A_6 and A_6A_7, and hence is not classified as a point at which L'

[6] In general, the curve L' obtained by removing a closed polygonal curve like BA_1A_2B from L will have at least one vertex less than L, since we must remove at least a triangle, thereby losing at least two vertices, while creating at most one new vertex.

intersects itself. Therefore, starting from A_6, we retrace the segment A_5A_6 in the reverse direction until we encounter either A_5 or A_7 (the vertices adjacent to A_6). In the case shown in Figure 13.4, the vertex $C = A_7$ is encountered first. The closed polygonal curve consisting of the segment $CA_6 \subset A_5A_6$ followed by the segment $A_6C \subset A_6A_7$ is a "2-gon", along which the integral of $f(z)$ vanishes (see Step 1). Therefore, removing the polygonal curve CA_6C from L', we obtain another closed polygonal curve

$$L'' = A_0BA_3A_4A_5CA_0$$

such that

$$\int_{L''} f(z)\,dz = \int_{L'} f(z)\,dz.$$

The curve L'' has one less vertex than L', since one vertex (A_6) is lost along with the 2-gon, and no new vertex replaces it, since $C = A_7$ (if $C = A_5$, we also lose one vertex, but if $C = A_5 = A_7$, we lose two vertices).

This argument is of a perfectly general character.[7] After a finite number of steps, we obtain either a closed polygonal curve which is a Jordan curve (L'' is such a curve in our example), or else we obtain a 2-gon. Thus, in either case, we find that $\int_L f(z)\,dz$ vanishes, thereby proving Cauchy's integral theorem for arbitrary closed polygonal curves. It should be noted that in extending the theorem from the case of triangles to the case of arbitrary closed polygonal curves, we used no function-theoretic considerations, but only geometric arguments of an elementary character.

STEP 6 (*Closed rectifiable curves*). Finally let L be an arbitrary closed rectifiable curve contained in the domain G. According to the key lemma (see p. 259), given any $\varepsilon > 0$, we can find a closed polygonal curve Λ inscribed in L and contained in G, such that

$$\left| \int_L f(z)\,dz - \int_\Lambda f(z)\,dz \right| < \varepsilon. \tag{13.16}$$

But as we have just seen,

$$\int_\Lambda f(z)\,dz = 0.$$

It follows that

$$\left| \int_L f(z)\,dz \right| < \varepsilon,$$

and since ε is arbitrarily small,

$$\int_L f(z)\,dz = 0.$$

The proof of Theorem 13.1 is now complete.

[7] The argument has already been encountered in the proof of Theorem 4.12.

In proving Theorem 13.1 we have assumed that G is simply connected. It is easy to see that this fact has only been used to guarantee that for sufficiently small ε, G contains the closed polygonal domains bounded by the segments of the polygonal curve Λ inscribed in L and satisfying (13.16). This observation leads to the following variants of Cauchy's integral theorem, where we drop the requirement that G itself be simply connected:

THEOREM 13.1′. *Let G be an arbitrary domain, and let $f(z)$ be a single-valued analytic function on G. Then*

$$\int_L f(z)\,dz = 0,$$

where L is any closed rectifiable curve contained in a simply connected subdomain of G.

THEOREM 13.1″. *Let G be an arbitrary domain, and let $f(z)$ be a single-valued analytic function on G. Then*

$$\int_L f(z)\,dz = 0,$$

where L is any closed rectifiable Jordan curve such that G contains both L and its interior $I(L)$.

Remark. The conditions given in Theorems 13.1′ and 13.1″ are sufficient but not necessary for $\int_L f(z)\,dz$ to vanish. For example, the functions $1/z$ and $1/z^2$ are both analytic on the domain G consisting of the whole plane with the single point $z = 0$ deleted, and G contains any circle $\gamma: |z| = r > 0$ but not $I(\gamma)$. However,

$$\int_\gamma \frac{dz}{z} = \int_0^{2\pi} \frac{ire^{i\theta}\,d\theta}{re^{i\theta}} = i\int_0^{2\pi} d\theta = 2\pi i,$$

(cf. Example 3, p. 252), while, on the other hand,

$$\int_\gamma \frac{dz}{z^2} = \int_0^{2\pi} \frac{ire^{i\theta}\,d\theta}{r^2e^{2i\theta}} = \frac{i}{r}\int_0^{2\pi} e^{-i\theta}\,d\theta = 0.$$

Furthermore, Figure 13.5 shows a doubly connected domain G and a curve $L \subset G$ such that

$$\int_L f(z)\,dz = 0,$$

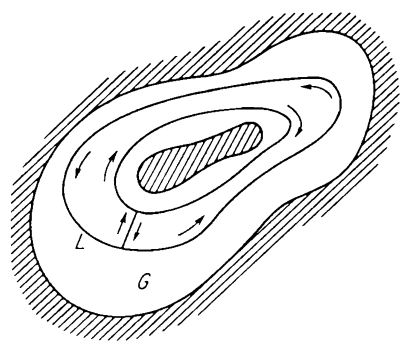

FIGURE 13.5

where $f(z)$ is an arbitrary single-valued analytic function on G, even though L is not contained in any simply connected subdomain of G.

In Theorem 13.1, it is assumed that $f(z)$ is analytic on a domain G containing the contour of integration of L. However, Theorem 13.1 can actually be extended to the case where L is the boundary of G:

THEOREM 13.3. *Let G be the interior of a closed rectifiable Jordan curve L, and let $f(z)$ be a single-valued function which is analytic on G and continuous on $\bar{G}$. Then*

$$\int_L f(z)\,dz = 0,$$

provided that G contains a point ζ such that every ray emanating from ζ intersects L in only one point.[8]

Proof. Suppose we write the equation of L in the form

$$z = \zeta + \lambda(\theta) \qquad (0 \leqslant \theta \leqslant 2\pi),$$

where θ is the polar angle of a system of polar coordinates with its origin at the point ζ, and $\lambda(2\pi) = \lambda(0)$. Let L_ρ be the curve with equation

$$z = \zeta + \rho\lambda(\theta) \qquad (0 \leqslant \theta \leqslant 2\pi).$$

Then L_ρ and L are homothetic figures with ray center ζ and ray ratio ρ (see p. 123). Moreover $L_\rho \subset G$ if $0 < \rho < 1$, and hence, according to Cauchy's integral theorem,

$$\int_{L_\rho} f(z)\,dz = 0 \qquad (0 < \rho < 1).$$

Now let $z_0, z_1, \ldots, z_n = z_0$ be the points of L corresponding to a partition $\mathscr{P} = \{\theta_0, \theta_1, \ldots, \theta_n\}$ of the interval $[0, 2\pi]$. Then the sum

$$\sum_{k=1}^{n} f(z_k)(z_k - z_{k-1}) = \sum_{k=1}^{n} f[\zeta + \lambda(\theta_k)][\lambda(\theta_k) - \lambda(\theta_{k-1})]$$

approaches $\int_L f(z)\,dz$ as $|\mathscr{P}|$ (the norm of $\mathscr{P}$) approaches zero. Since L_ρ and L are homothetic figures, the sum

$$\sum_{k=1}^{n} f[\zeta + \rho\lambda(\theta_k)][\rho\lambda(\theta_k) - \rho\lambda(\theta_{k-1})]$$

approaches $\int_{L_\rho} f(z)\,dz$ as $|\mathscr{P}| \to 0$. It follows that

$$\int_L \rho f[\zeta + \rho\lambda(\theta)]\,dz = \int_{L_\rho} f(z)\,dz = 0.$$

---

[8] Considerations which will not be given here show that this proviso is actually unnecessary.

Next, using this result, we write

$$\int_L f(z)\,dz = \int_L f[\zeta + \lambda(\theta)]\,dz = \int_L \{f[\zeta + \lambda(\theta)] - \rho f[\zeta + \rho\lambda(\theta)]\}\,dz$$
$$= (1 - \rho)\int_L f[\zeta + \lambda(\theta)]\,dz + \rho\int_L \{f[\zeta + \lambda(\theta)] - f[\zeta + \rho\lambda(\theta)]\}\,dz. \tag{13.17}$$

Since $f(z)$ is uniformly continuous on $\bar{G}$ (see Theorem 3.17), given any $\varepsilon > 0$, there is a $\delta = \delta(\varepsilon) > 0$ such that

$$|f(z) - f(z')| < \varepsilon$$

for every pair of points z, z' in $\bar{G}$ such that $|z - z'| < \delta$. Choosing

$$1 - \rho < \frac{\delta}{m},$$

where

$$m = \max_{0 \leq \theta \leq 2\pi} |\lambda(\theta)|,$$

we have

$$|f[\zeta + \lambda(\theta)] - f[\zeta + \rho\lambda(\theta)]| < \varepsilon,$$

and hence, according to (13.17),

$$\left|\int_L f(z)\,dz\right| < [(1 - \rho)M + \rho\varepsilon]l$$

where l is the length of L, and

$$M = \max_{z \in L} |f(z)|.$$

As $\rho \to 1$, the last inequality becomes

$$\left|\int_L f(z)\,dz\right| < \varepsilon l,$$

so that letting $\varepsilon \to 0$, we obtain

$$\int_L f(z)\,dz = 0,$$

as asserted.

Remark. In particular, L can be a circle or the boundary of any convex polygon. Moreover, since Theorem 13.3 holds for any triangle, it holds for any closed polygonal Jordan curve (convex or not). However, this does not mean that we can use the lemma of Sec. 64 to extend Theorem 13.3 to the case of an arbitrary closed rectifiable Jordan curve L, since a polygonal curve Λ inscribed in L may partially lie outside $\bar{G} = \overline{I(L)}$, in which case

$$\int_\Lambda f(z)\,dz$$

is meaningless.

66. Application to the Evaluation of Definite Integrals

In early researches, Cauchy used his integral theorem as a tool for evaluating various definite integrals of functions of a real variable, especially improper integrals. This application of Theorem 13.1, which played a key role in the history of the subject, is illustrated by the following examples:

Example 1. *Evaluate the Fresnel integrals*

$$\int_0^\infty \cos x^2\, dx, \qquad \int_0^\infty \sin x^2\, dx, \tag{13.18}$$

encountered in the theory of diffraction.

We begin by deriving two preliminary results which will be needed to calculate (13.18), i.e., the formula

$$\int_0^\infty e^{-x^2} = \frac{\sqrt{\pi}}{2} \tag{13.19}$$

and the inequality

$$\sin\theta \geqslant \frac{2}{\pi}\theta \qquad \left(0 \leqslant \theta \leqslant \frac{\pi}{2}\right). \tag{13.20}$$

To prove (13.19), let Q be the square $-R \leqslant x \leqslant R$, $-R \leqslant y \leqslant R$, let $C_1: |z| = R$ be the circle inscribed in Q, and let $C_2: |z| = \sqrt{2}\,R$ be the circle circumscribed about Q. Then, if $\bar{K}_1$ and $\bar{K}_2$ are the closed disks with boundaries C_1 and C_2, respectively, we have

$$\begin{aligned} \iint_{\bar{K}_1} e^{-x^2-y^2}\, dx\, dy &< \iint_Q e^{-x^2-y^2}\, dx\, dy \\ &= \left(\int_{-R}^{R} e^{-x^2}\, dx\right)^2 < \iint_{\bar{K}_2} e^{-x^2-y^2}\, dx\, dy, \end{aligned} \tag{13.21}$$

since the exponential is positive. If we introduce polar coordinates r and θ, (13.21) becomes

$$\int_0^{2\pi}\int_0^R e^{-r^2}\, r\, dr\, d\theta < 4\left(\int_0^R e^{-x^2}\, dx\right)^2 < \int_0^{2\pi}\int_0^{\sqrt{2}R} e^{-r^2}\, r\, dr\, d\theta.$$

Evaluating the integrals, and taking the square roots of all terms of the resulting inequalities, we find that

$$\sqrt{\pi(1 - e^{-R^2})} < 2\int_0^R e^{-x^2}\, dx < \sqrt{\pi(1 - e^{-2R^2})}.$$

Thus

$$\lim_{R\to\infty} 2\int_0^R e^{-x^2}\, dx = \sqrt{\pi},$$

which implies (13.19).

To prove (13.20), we note that

$$f(\theta) = \frac{\sin \theta}{\theta}$$

is a decreasing function on the interval $(0, \pi/2)$, since

$$f'(\theta) = \frac{\theta \cos \theta - \sin \theta}{\theta^2} = \frac{\cos \theta (\theta - \tan \theta)}{\theta^2} < 0$$

if $0 < \theta < \pi/2$. Therefore $f(\theta) > f(\pi/2)$ if $0 < \theta < \pi/2$, i.e.,

$$\frac{\sin \theta}{\theta} > \frac{2}{\pi}$$

or

$$\sin \theta > \frac{2}{\pi}\theta \qquad \left(0 < \theta < \frac{\pi}{2}\right).$$

Moreover, it is obvious that $\sin \theta = 2\theta/\pi$ if $\theta = 0$ or $\theta = \pi/2$, and hence (13.20) is proved.

Turning to the evaluation of the Fresnel integrals (13.18), we consider the function

$$f(z) = e^{iz^2}$$

(the reason for this choice will soon be apparent). Since both e^z and iz^2 are entire functions, it follows from the rule for differentiating a composite function (see p. 108) that $f(z)$ is entire, with derivative $f'(z) = 2ize^{iz^2}$. Therefore we can apply Cauchy's theorem to $f(z)$, with any closed rectifiable curve as the contour L. For L we choose the contour shown in Figure 13.6, consisting of the segment $OA: 0 \leqslant x \leqslant R$ of the nonnegative real axis, the arc AB of the circle of radius R with center at the origin, and the segment BO of the line bisecting the angle formed by the nonnegative real and imaginary axes. According to Cauchy's integral theorem,

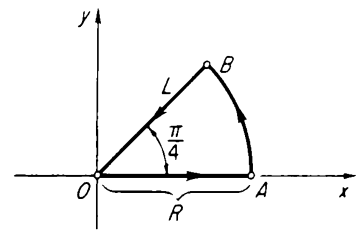

FIGURE 13.6

$$\int_L e^{iz^2}\,dz = \int_{OA} e^{iz^2}\,dz + \int_{AB} e^{iz^2}\,dz + \int_{BO} e^{iz^2}\,dz = 0. \qquad (13.22)$$

On the segment OA, z reduces to a real number x, so that $dz = dx$, and

$$J_1(R) = \int_{OA} e^{iz^2}\,dz = \int_0^R e^{ix^2}\,dx.$$

On the arc AB

$$z = Re^{i\theta} \quad (0 \leqslant \theta \leqslant \pi/4), \qquad z^2 = R^2 e^{2i\theta}, \qquad dz = iRe^{i\theta}\,d\theta,$$

so that

$$J_2(R) = \int_{AB} e^{iz^2}\,dz = \int_0^{\pi/4} \exp\,(iR^2e^{2i\theta})\,iRe^{i\theta}\,d\theta.$$

Finally, on the segment BO

$$z = re^{i\pi/4} \quad (R \geqslant r \geqslant 0), \quad z^2 = r^2 e^{i\pi/2} = ir^2, \quad dz = e^{i\pi/4}\, dr,$$

and consequently

$$J_3(R) = \int_{BO} e^{iz^2}\, dz = \int_R^0 e^{-r^2} e^{i\pi/4}\, dr = -e^{i\pi/4} \int_0^R e^{-r^2}\, dr$$
$$= -\frac{\sqrt{2}}{2}(1 + i) \int_0^R e^{-r^2}\, dr.$$

Next let $R \to \infty$. Then $J_3(R)$ approaches the limit

$$-\frac{\sqrt{2}}{2}(1 + i) \int_0^\infty e^{-r^2}\, dr = -\frac{\sqrt{2\pi}}{4}(1 + i),$$

where we have used formula (13.19). Moreover, $J_2(R) \to 0$ as $R \to \infty$. To show this, we estimate the modulus of $J_2(R)$, obtaining

$$|J_2(R)| \leqslant R \int_0^{\pi/4} |\exp [iR^2(\cos 2\theta + i \sin 2\theta)]|\, |e^{i\theta}|\, d\theta.$$

Since $|e^{i\theta}| = 1$ and

$$|\exp [iR^2(\cos 2\theta + i \sin 2\theta)]| = \exp (-R^2 \sin 2\theta),$$

we have

$$|J_2(R)| \leqslant R \int_0^{\pi/4} \exp (-R^2 \sin 2\theta)\, d\theta.$$

But according to (13.20), $\sin 2\theta \geqslant 4\theta/\pi$ if $0 \leqslant 2\theta \leqslant \pi/2$. Therefore

$$|J_2(R)| \leqslant R \int_0^{\pi/4} \exp (-4R^2\theta/\pi)\, d\theta = R \frac{\exp (-4R^2\theta/\pi)}{-4R^2/\pi} \Bigg|_{\theta=0}^{\theta=\pi/4}$$
$$= \frac{\pi}{4} \frac{1 - e^{-R^2}}{R},$$

so that

$$\lim_{R \to \infty} J_2(R) = 0,$$

as asserted. Finally, we consider the integral

$$J_1(R) = \int_0^R e^{ix^2}\, dx = \int_0^R \cos x^2\, dx + i \int_0^R \sin x^2\, dx.$$

Since

$$J_1(R) + J_2(R) + J_3(R) = 0$$

[cf. (13.22)], it follows that

$$\lim_{R \to \infty} J_1(R) = -\lim_{R \to \infty} J_2(R) - \lim_{R \to \infty} J_3(R) = \frac{\sqrt{2\pi}}{4}(1 + i),$$

i.e.,

$$\lim_{R \to \infty} \left[\int_0^R \cos x^2\, dx + i \int_0^R \sin x^2\, dx \right] = \frac{\sqrt{2\pi}}{4}(1 + i),$$

which implies that the improper integrals

$$\int_0^\infty \cos x^2\,dx = \lim_{R\to\infty}\int_0^R \cos x^2\,dx$$

and

$$\int_0^\infty \sin x^2\,dx = \lim_{R\to\infty}\int_0^R \sin x^2\,dx$$

exist, and have the same value

$$\int_0^\infty \cos x^2\,dx = \int_0^\infty \sin x^2\,dx = \frac{\sqrt{2\pi}}{4}.$$

Example 2. *Evaluate the integral*

$$\int_{-\infty}^{\infty} e^{-\lambda x^2}\cos(2\alpha\lambda x)\,dx \qquad (\lambda > 0,\ \alpha > 0). \tag{13.23}$$

To calculate (13.23), we integrate the function

$$f(z) = e^{-\lambda z^2}$$

along the rectangular contour L shown in Figure 13.7. Since $f(z)$ is entire,

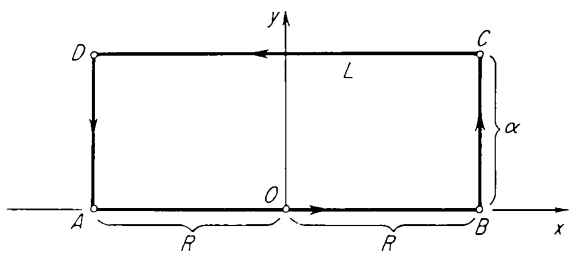

FIGURE 13.7

it meets the conditions of Theorem 13.1 with any closed rectifiable curve as the contour L, and therefore

$$\int_L e^{-\lambda z^2}\,dz = \int_{AB} e^{-\lambda z^2}\,dz + \int_{BC} e^{-\lambda z^2}\,dz + \int_{CD} e^{-\lambda z^2}\,dz + \int_{DA} e^{-\lambda z^2}\,dz = 0. \tag{13.24}$$

On AB

$$z = x \quad (-R \leqslant x \leqslant R), \qquad dz = dx,$$

so that

$$J_1(R) = \int_{AB} e^{-\lambda z^2}\,dz = \int_{-R}^{R} e^{-\lambda x^2}\,dx.$$

On BC

$$z = R + iy \quad (0 \leqslant y \leqslant \alpha), \qquad z^2 = R^2 + 2iRy - y^2, \qquad dz = i\,dy,$$

and consequently

$$J_2(R) = \int_{BC} e^{-\lambda z^2}\,dz = \int_0^{\alpha} \exp\,[-\lambda(R^2 + 2iRy - y^2]i\,dy$$
$$= i\int_0^{\alpha} \exp\,[-\lambda(R^2 - y^2)]\exp\,(-2iR\lambda y)\,dy.$$

On CD

$$z = x + i\alpha \quad (R \geqslant x \geqslant -R), \qquad z^2 = x^2 + 2i\alpha x - \alpha^2, \qquad dz = dx,$$

so that

$$J_3(R) = \int_{CD} e^{-\lambda z^2}\,dz = \int_R^{-R} \exp\,[-\lambda(x^2 + 2i\alpha x - \alpha^2)]\,dx$$
$$= -e^{\lambda\alpha^2}\int_{-R}^{R} \exp\,(-\lambda x^2 - 2i\alpha\lambda x)\,dx$$
$$= -e^{\lambda\alpha^2}\int_{-R}^{R} e^{-\lambda x^2}\,[\cos\,(2\alpha\lambda x) - i\sin\,(2\alpha\lambda x)]\,dx.$$

Finally, on DA

$$z = -R + iy \quad (\alpha \geqslant y \geqslant 0), \qquad z^2 = R^2 - 2iRy - y^2, \qquad dz = i\,dy,$$

and consequently

$$J_4(R) = \int_{DA} e^{-\lambda z^2}\,dz = \int_{\alpha}^{0} \exp\,[-\lambda(R^2 - 2iRy - y^2)]i\,dy$$
$$= -i\int_0^{\alpha} \exp\,[-\lambda(R^2 - y^2)]\exp\,(2iR\lambda y)\,dy.$$

Next let $R \to \infty$. Then the integrals $J_2(R)$ and $J_4(R)$ approach zero. In fact,

$$|J_2(R)| \leqslant \int_0^{\alpha} |\exp\,[-\lambda(R^2 - y^2)]|\,|\exp\,(-2iR\lambda y)|\,dy$$
$$= \int_0^{\alpha} \exp\,[-\lambda(R^2 - y^2)]\,dy,$$

so that for $R > \alpha$

$$|J_2(R)| \leqslant \int_0^{\alpha} \exp\,[-\lambda(R^2 - \alpha^2)]\,dy = \alpha\exp\,[-\lambda(R^2 - \alpha^2)] \to 0$$

as $R \to \infty$. Similarly,

$$|J_4(R)| \leqslant \int_0^{\alpha} |\exp\,[-\lambda(R^2 - y^2)]|\,|\exp\,(2iR\lambda y)|\,dy$$
$$= \int_0^{\alpha} \exp\,[-\lambda(R^2 - y^2)]\,dy \to 0$$

as $R \to \infty$. Moreover,

$$\lim_{R\to\infty} J_1(R) = \int_{-\infty}^{\infty} e^{-\lambda x^2}\,dx = \frac{1}{\sqrt{\lambda}}\int_{-\infty}^{\infty} e^{-(\sqrt{\lambda}x)^2}\,d(\sqrt{\lambda}x) = \sqrt{\frac{\pi}{\lambda}}$$

[cf. (13.19)]. Finally, since

$$J_1(R) + J_2(R) + J_3(R) + J_4(R) = 0$$

[cf. (13.24)], we have

$$e^{\lambda\alpha^2}\int_{-\infty}^{\infty} e^{-\lambda x^2}\left[\cos(2\alpha\lambda x) - i\sin(2\alpha\lambda x)\right] = -\lim_{R\to\infty} J_3(R)$$

$$= \lim_{R\to\infty} J_1(R) + \lim_{R\to\infty} J_2(R) + \lim_{R\to\infty} J_4(R) = \sqrt{\frac{\pi}{\lambda}}.$$

Taking the real part of both sides of this formula, we find that

$$e^{\lambda\alpha^2}\int_{-\infty}^{\infty} e^{-\lambda x^2}\cos(2\alpha\lambda x)\,dx = \sqrt{\frac{\pi}{\lambda}},$$

or

$$\int_{-\infty}^{\infty} e^{-\lambda x^2}\cos(2\alpha\lambda x)\,dx = \sqrt{\frac{\pi}{\lambda}}\,e^{-\lambda\alpha^2}.$$

Example 3. *Evaluate the integral*

$$\int_0^{\infty}\frac{\sin x}{x}\,dx. \qquad (13.25)$$

To calculate (13.25), we consider the function

$$f(z) = \frac{e^{iz}}{z},$$

which is defined and differentiable [with derivative $f'(z) = e^{iz}(iz - 1)/z^2$] on the domain G consisting of the whole z-plane with the single point $z = 0$ deleted. Let L be the contour shown in Figure 13.8. Since G contains L and its interior $I(L)$, it follows from Theorem 13.1″, p. 269 that

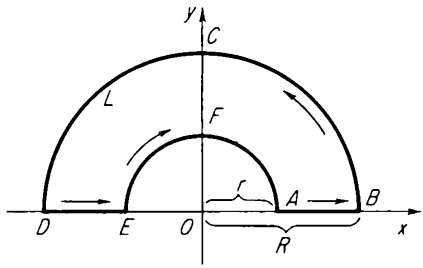

FIGURE 13.8

$$\int_L \frac{e^{iz}}{z}\,dz = \int_{AB}\frac{e^{iz}}{z}\,dz + \int_{BCD}\frac{e^{iz}}{z}\,dz + \int_{DE}\frac{e^{iz}}{z}\,dz + \int_{EFA}\frac{e^{iz}}{z}\,dz = 0. \qquad (13.26)$$

On AB

$$z = x \quad (r \leqslant x \leqslant R), \qquad dz = dx,$$

so that

$$J_1(r, R) = \int_{AB}\frac{e^{iz}}{z}\,dz = \int_r^R \frac{e^{ix}}{x}\,dx = \int_r^R \frac{\cos x}{x}\,dx + i\int_r^R \frac{\sin x}{x}\,dx. \qquad (13.27)$$

On BCD

$$z = Re^{i\theta} \quad (0 \leqslant \theta \leqslant \pi), \quad dz = iRe^{i\theta}\,d\theta,$$

and consequently

$$J_2(R) = \int_{BCD} \frac{e^{iz}}{z}\,dz = \int_0^{\pi} \frac{\exp(iRe^{i\theta})}{Re^{i\theta}}\, iRe^{i\theta}\,d\theta$$
$$= i\int_0^{\pi} \exp(iR\cos\theta - R\sin\theta)\,d\theta.$$

On DE

$$z = x \quad (-R \geqslant x \geqslant -r), \qquad dz = dx,$$

so that

$$J_3(r, R) = \int_{DE} \frac{e^{iz}}{z}\,dz = \int_{-R}^{-r} \frac{e^{ix}}{x}\,dx = \int_{-R}^{-r} \frac{\cos x}{x}\,dx + i\int_{-R}^{-r} \frac{\sin x}{x}\,dx. \quad (13.28)$$

If we change the variable of integration from x to $-x$, (13.28) becomes

$$J_3(r, R) = -\int_r^R \frac{\cos x}{x}\,dx + i\int_r^R \frac{\sin x}{x}\,dx. \quad (13.29)$$

Finally, on EFA

$$z = re^{i\theta} \quad (\pi \geqslant \theta \geqslant 0), \qquad dz = ire^{i\theta}\,d\theta,$$

and consequently

$$J_4(r) = \int_{EFA} \frac{e^{iz}}{z}\,dz = \int_{\pi}^{0} \frac{\exp(ire^{i\theta})}{re^{i\theta}}\, ire^{i\theta}\,d\theta$$
$$= -i\int_0^{\pi} \exp(ir\cos\theta - r\sin\theta)\,d\theta.$$

Now let $R \to \infty$. Since

$$|J_2(R)| \leqslant \int_0^{\pi} |\exp(iR\cos\theta)|\,|\exp(-R\sin\theta)|\,d\theta$$
$$= \int_0^{\pi} \exp(-R\sin\theta)\,d\theta$$
$$< \int_0^{\pi} \exp(-2R\theta/\pi)\,d\theta$$
$$= \frac{\exp(-2R\theta/\pi)}{-2R/\pi}\Bigg|_{\theta=0}^{\theta=\pi} = \frac{\pi}{2}\,\frac{1 - e^{-2R}}{R} < \frac{\pi}{2R}$$

[cf. (13.20)], we have

$$\lim_{R\to\infty} J_2(R) = 0.$$

Next let $r \to 0$ and calculate

$$\lim_{r\to\infty} J_4(r).$$

Since the function e^{iz} is continuous at the point $z = 0$, where it takes the value 1, then, given any $\varepsilon > 0$, we can find a $\delta = \delta(\varepsilon) > 0$ such that

$$|e^{iz} - 1| = |\exp(ir\cos\theta - r\sin\theta) - 1| < \varepsilon$$

if $|z| = r < \delta$. It follows that

$$\left| J_4(r) - \left(-i \int_0^{\pi} 1 \cdot d0 \right) \right| = \left| -i \int_0^{\pi} [\exp(ir \cos 0 - r \sin 0) - 1]\, d0 \right| < \pi\varepsilon,$$

so that

$$\lim_{r \to 0} J_4(r) = -i \int_0^{\pi} d0 = -\pi i.$$

Finally, since

$$J_1(r, R) + J_2(R) + J_3(r, R) + J_4(r) = 0$$

[cf. (13.26)], it follows that

$$\lim_{\substack{r \to 0 \\ R \to \infty}} [J_1(r, R) + J_3(r, R)] = \lim_{\substack{r \to 0 \\ R \to \infty}} [-J_2(R) - J_4(r)] = \pi i.$$

Therefore, using (13.27) and (13.29), we see that the improper integral

$$\frac{1}{2i} \lim_{\substack{r \to 0 \\ R \to \infty}} [J_1(r, R) + J_3(r, R)] = \lim_{\substack{r \to 0 \\ R \to \infty}} \int_r^R \frac{\sin x}{x}\, dx$$

exists and has the value

$$\int_0^{\infty} \frac{\sin x}{x}\, dx = \frac{\pi}{2}.$$

67. Cauchy's Integral Theorem for a System of Contours

We now prove a result which will be used repeatedly later on:

THEOREM 13.4 (*Cauchy's integral theorem for a system of contours*). *Let G be an arbitrary domain, and let $f(z)$ be a single-valued analytic function on G. Moreover, let $\Gamma, \gamma_1, \ldots, \gamma_n$ be a system of $n + 1$ closed rectifiable Jordan curves contained in G which satisfy the following conditions*:

1. *$I(\Gamma)$ contains every curve $\gamma_1, \ldots, \gamma_n$;*
2. *For every $k = 1, \ldots, n$, $E(\gamma_k)$ contains γ_j $(j \neq k)$;*
3. *G contains the multiply connected domain*

$$D = I(\Gamma) - \overline{I(\gamma_1)} - \cdots - \overline{I(\gamma_n)},$$

with boundary $\Gamma \cup \gamma_1 \cup \cdots \cup \gamma_n$.

Then

$$\int_{\Gamma} f(z)\, dz = \int_{\gamma_1} f(z)\, dz + \cdots + \int_{\gamma_n} f(z)\, dz, \qquad (13.30)$$

where all the integrals are taken in the same direction, e.g., in such a way that the interiors of the curves $\Gamma, \gamma_1, \ldots, \gamma_n$ are to the left of an observer moving along the curves.[9]

Proof. First we note that the theorem is trivial in the special case where G is simply connected, since then all the integrals in both sides of

[9] This will be regarded as the positive direction of traversing the system of contours.

(13.30) vanish. To prove the theorem in the general case, we draw $n + 1$ rectifiable Jordan arcs $\delta_1', \delta_2', \ldots, \delta_{n+1}'$ contained in G such that δ_1' joins a point $z_0 \in \Gamma$ to a point $\zeta_1 \in \gamma_1$, δ_2' joins a point $z_1 \in \gamma_1$ ($z_1 \neq \zeta_1$) to a point $\zeta_2 \in \gamma_2$, and so on, until finally δ_{n+1}' joins a point $z_n \in \gamma_n$ to a point $\zeta_0 \in \Gamma$ ($\zeta_0 \neq z_0$). In general, these arcs may partially leave the domain D, but they can always be replaced by $n + 1$ new arcs $\delta_1, \delta_2, \ldots, \delta_{n+1}$ which, except for their end points, are contained in D. In fact, as we go along δ_1' from the point $z_0 \in \Gamma$ to the point $\zeta_1 \in \gamma_1$, we need only

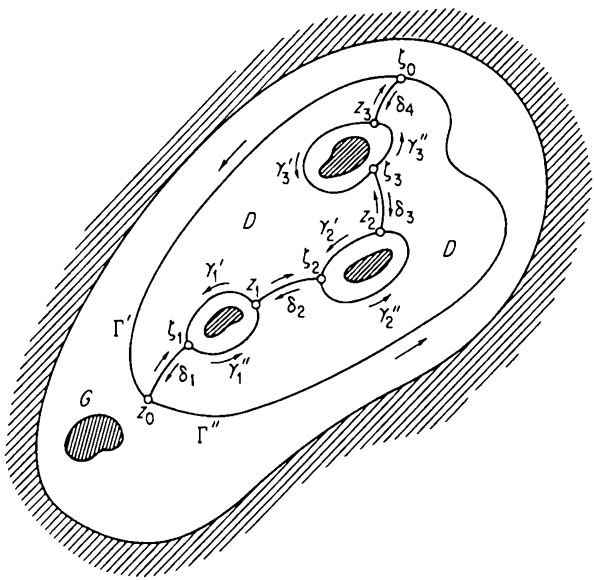

FIGURE 13.9

note the last point of intersection of δ_1' with Γ and the first point of intersection of δ_1' with γ_1. Then the part of δ_1' lying between these two points of intersection gives the required arc δ_1, and similarly we can construct the rest of the arcs $\delta_2, \ldots, \delta_{n+1}$.[10] The initial and final points of the arcs $\delta_1, \delta_2, \ldots, \delta_n$ divide each of the closed curves $\Gamma, \gamma_1, \ldots, \gamma_n$ into two arcs, which we indicate by adding one or two primes to the symbols denoting the original curves:[11]

$$\Gamma = \Gamma' - \Gamma'', \qquad \gamma_1 = \gamma_1' - \gamma_1'', \ldots, \gamma_n = \gamma_n' - \gamma_n''.$$

[10] In general, the arcs $\delta_1, \delta_2, \ldots, \delta_n$ so obtained have points of intersection, but it can be shown (cf. Problem 13.14) that they can always be replaced by *disjoint* arcs (which we denote by the same symbols) satisfying the same conditions.

[11] By $\gamma - \delta$ is meant the curve $\gamma + (-\delta)$, where γ and δ are two curves (arcs) with a common final point.

As the common initial point of the arcs Γ', Γ'' we choose the initial point z_0 of the arc δ_1, and as the common final point of Γ', Γ'' we choose the final point ζ_0 of δ_{n+1}. Similarly, the arcs γ_1', γ_1'' have initial point ζ_1 (the final point of δ_1) and final point z_1 (the initial point of δ_2) and so on, with ζ_n (the final point of δ_n) and z_n (the initial point of δ_{n+1}) as the initial and final points, respectively, of the arcs γ_n', γ_n''. The whole construction is illustrated by Figure 13.9, which corresponds to the case $n = 3$.

Now consider the two closed rectifiable Jordan curves

$$L' = \Gamma' - \delta_{n+1} - \gamma_n' - \cdots - \gamma_1' - \delta_1,$$

and

$$L'' = -\Gamma'' + \delta_1 + \gamma_1'' + \cdots + \gamma_n'' + \delta_{n+1}.$$

Clearly, L' and L'' are both traversed in the same (positive) direction, and

$$L' \subset G, \quad L'' \subset G, \quad I(L') \subset D \subset G, \quad I(L'') \subset D \subset G.$$

Therefore, by Theorem 13.1,

$$\int_{L'} f(z)\,dz = \int_{L''} f(z)\,dz = 0,$$

or

$$\begin{aligned} 0 &= \left(\int_{L'} + \int_{L''}\right) f(z)\,dz \\ &= \left(\int_{\Gamma'} + \int_{-\Gamma''} + \int_{\delta_1} + \int_{-\delta_1} + \int_{-\gamma_1'} + \int_{\gamma_1''} + \cdots \right. \\ &\qquad \left. + \int_{-\gamma_n'} + \int_{\gamma_n''} + \int_{\delta_{n+1}} + \int_{-\delta_{n+1}}\right) f(z)\,dz. \end{aligned} \tag{13.31}$$

But it is easy to see that

$$\left(\int_{\delta_j} + \int_{-\delta_j}\right) f(z)\,dz = 0 \qquad (j = 1, 2, \ldots, n+1),$$

and

$$\left(\int_{\Gamma'} + \int_{-\Gamma''}\right) f(z)\,dz = \int_{\Gamma} f(z)\,dz,$$

$$\left(\int_{-\gamma_k'} + \int_{\gamma_k''}\right) f(z)\,dz = \int_{-\gamma_k} f(z)\,dz \qquad (k = 1, 2, \ldots, n).$$

Therefore (13.31) implies (13.30), and the proof of Theorem 13.4 is complete.

68. Path-Independent Integrals. Primitives

Let G be a simply connected domain, and let $f(z)$ be a single-valued analytic function on G. Suppose z_0 and z are any two points of G, while L_1

and L_2 are any two rectifiable curves joining z_0 to z and contained in G. Then it is an immediate consequence of Cauchy's integral theorem that

$$\int_{L_1} f(z)\,dz = \int_{L_2} f(z)\,dz, \tag{13.32}$$

since

$$0 = \int_{L_1 - L_2} f(z)\,dz = \int_{L_1} f(z)\,dz - \int_{L_2} f(z)\,dz.$$

Therefore the integral of the function $f(z)$ along any rectifiable curve joining z_0 to z is *path-independent*, i.e., depends only on its end points z_0 and z, a fact which can be emphasized by writing (13.32) as

$$\int_{z_0}^{z} f(\zeta)\,d\zeta.$$

If the lower limit of integration z_0 is held fixed, while the upper limit of integration z is allowed to vary, we obtain a single-valued function

$$F(z) = \int_{z_0}^{z} f(\zeta)\,d\zeta \qquad (z_0, z \in G). \tag{13.33}$$

We now show that $F(z)$ is itself an analytic function on G, with derivative $f(z)$.

THEOREM 13.5. *Let G be a simply connected domain, and let $f(z)$ be a single-valued analytic function on G. Then the single-valued function $F(z)$ defined by* (13.33) *is analytic on G and*

$$F'(z) = f(z).$$

Proof. Given a point $z \in G$, let the modulus of Δz be so small that G contains $\overline{I(\gamma)}$, where γ is the circle with center z and radius $|\Delta z|$. Then

$$\Delta F \equiv \int_{z_0}^{z+\Delta z} f(\zeta)\,d\zeta - \int_{z_0}^{z} f(\zeta)\,d\zeta = \int_{z}^{z+\Delta z} f(\zeta)\,d\zeta,$$

where the last integral can be evaluated along the line segment joining z to $z + \Delta z$. It follows that

$$\begin{aligned}\Delta F &= \int_{z}^{z+\Delta z} f(z)\,d\zeta + \int_{z}^{z+\Delta z} [f(\zeta) - f(z)]\,d\zeta \\ &= \Delta z f(z) + \int_{z}^{z+\Delta z} [f(\zeta) - f(z)]\,d\zeta,\end{aligned}$$

and hence

$$\frac{\Delta F}{\Delta z} = f(z) + \frac{1}{\Delta z}\int_{z}^{z+\Delta z} [f(\zeta) - f(z)]\,d\zeta$$

or

$$\left|\frac{\Delta F}{\Delta z} - f(z)\right| \leqslant \frac{1}{|\Delta z|}\left|\int_{z}^{z+\Delta z} [f(\zeta) - f(z)]\,d\zeta\right| \leqslant \frac{1}{|\Delta z|} M(\Delta z)|\Delta z| = M(\Delta z),$$

where

$$M(\Delta z) = \max_{|\zeta - z| \leq |\Delta z|} |f(\zeta) - f(z)|.$$

[The existence of $M(\Delta z)$ follows from Theorem 3.16 and the continuity of $f(\zeta) - f(z)$ on the disk $\overline{I(\gamma)}$.] Since $f(\zeta)$ is continuous at z,

$$\lim_{\Delta z \to 0} \left| \frac{\Delta F}{\Delta z} - f(z) \right| = \lim_{\Delta z \to 0} M(\Delta z) = 0,$$

i.e.,

$$\lim_{\Delta z \to 0} \frac{\Delta F}{\Delta z} \equiv F'(z) = f(z),$$

as asserted.

Let $f(z)$ be a single-valued analytic function on a domain G. Then a function $\Phi(z)$ is said to be a *primitive* (synonymously, an *indefinite integral* or *anti-derivative*) of $f(z)$ on G if $\Phi(z)$ is single-valued and analytic on G, and

$$\Phi'(z) = f(z) \qquad (z \in G).$$

Thus, according to Theorem 13.5, the function (13.33) is a primitive of $f(z)$ on G. We now show that (13.33) is essentially the most general primitive of $f(z)$ on G, if G is simply connected.

THEOREM 13.6. *Let G be a simply connected domain, and let $f(z)$ be a single-valued analytic function on G. Then any primitive of $f(z)$ on G can be represented in the form*

$$\Phi(z) = \int_{z_0}^{z} f(\zeta)\, d\zeta + C, \tag{13.34}$$

where $z_0 \in G$ and C is a constant.

Proof. Writing

$$\varphi(z) = \Phi(z) - \int_{z_0}^{z} f(\zeta)\, d\zeta = u(x, y) + iv(x, y),$$

we have

$$\varphi'(z) = \Phi'(z) - f(z) = 0 \qquad (z \in G).$$

But according to Theorem 7.2,

$$\varphi'(z) = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i \frac{\partial u}{\partial y},$$

so that

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = \frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} = 0$$

everywhere in G. Since $u(x, y)$ and $v(x, y)$ are differentiable functions of x and y, it follows that

$$u(x, y) = C_1, \qquad v(x, y) = C_2,$$

where C_1 and C_2 are real constants. Therefore

$$\varphi(z) = C_1 + iC_2 = C \qquad (z \in G),$$

which implies (13.34).

COROLLARY (*Fundamental theorem of integral calculus for complex functions*). *Let G be a simply connected domain, and let $f(z)$ be a single-valued analytic function on G. Then if $z_0, z \in G$,*

$$\int_{z_0}^{z} f(\zeta)\, d\zeta = \Phi(z) - \Phi(z_0),$$

where $\Phi(z)$ is any primitive of $f(z)$ on G.

Proof. We merely use the fact that

$$\Phi(z_0) = C.$$

Remark. According to the corollary to Theorem 13.6, we can calculate integrals of elementary functions of a complex variable by using formulas resembling those for elementary functions of a real variable. For example,

$$\int_{z_0}^{z} \zeta^n\, d\zeta = \frac{z^{n+1} - z_0^{n+1}}{n+1}$$

if n is an integer other than -1, and

$$\int_{z_0}^{z} e^{\zeta}\, d\zeta = e^z - e^{z_0},$$

$$\int_{z_0}^{z} \cos\zeta\, d\zeta = \sin z - \sin z_0,$$

$$\int_{z_0}^{z} \sin\zeta\, d\zeta = \cos z_0 - \cos z.$$

69. The Integral as a Function of its Upper Limit in a Multiply Connected Domain

Let G be a multiply connected domain, and let $f(z)$ be a single-valued analytic function on G. If

$$\int_L f(z)\, dz = 0$$

for every closed rectifiable curve contained in G, then all the considerations of the preceding section apply to the integral

$$F(z) = \int_{z_0}^{z} f(\zeta)\, d\zeta \qquad (z_0, z \in G), \tag{13.35}$$

and we again find that $F(z)$ is a single-valued analytic function on G, in fact, a primitive of $f(z)$ on G. For example, this is the case if

$$f(z) = \frac{1}{z^2 + 1} \tag{13.36}$$

and G is the ring-shaped domain or *annulus* $1 < |z| < R$, where $R > 1$ is arbitrary (see Problem 13.16). However, suppose G contains at least one closed rectifiable curve $L = BCDE$ such that[12]

$$\int_L f(z)\, dz \neq 0,$$

and let $A = z_0$, $F = z$ be arbitrary points of G (see Figure 13.10). Then

$$\begin{aligned}\int_{ABCDF} f(z)\, dz - \int_{ABEDF} f(z)\, dz &= \int_{BCD} f(z)\, dz - \int_{BED} f(z)\, dz \\ &= \int_L f(z)\, dz \neq 0,\end{aligned}$$

and therefore

$$\int_{z_0}^{z} f(\zeta)\, d\zeta$$

has at least two different values at z, i.e., we can continue to write (13.35), but now the function $F(z)$ is multiple-valued on G. The appropriate generalization of Theorem 13.5 to the case of multiply connected domains is given by

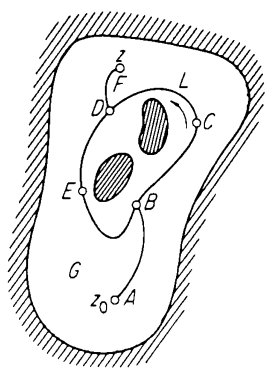

FIGURE 13.10

THEOREM 13.7. *Let G be a multiply connected domain, and let $f(z)$ be a single-valued analytic function on G. Then the function* (13.35), *which is in general multiple-valued, has single-valued analytic branches on any simply connected subdomain $D \subset G$, and*

$$F'(z) = f(z) \qquad (z \in D),$$

where $F'(z)$ denotes the derivative of any of these single-valued branches.

Proof. If G contains no closed rectifiable curve L such that

$$\int_L f(z)\, dz \neq 0,$$

[12] For example, suppose $f(z) = 1/z$ and G is the domain consisting of all points $z \neq 0$, or suppose $f(z)$ is given by (13.36) and G is the set of all points z satisfying $z^2 + 1 \neq 0$ (see Problem 13.16).

then $F(z)$ is single-valued and the proof of the theorem is identical with that of Theorem 13.5. Otherwise, let

$$F(z) = \int_{z_0}^{z_1} f(\zeta)\, d\zeta + \int_{z_1}^{z} f(\zeta)\, d\zeta \equiv \int_{z_0}^{z_1} f(\zeta)\, d\zeta + \varphi(z), \quad (13.37)$$

where z_1 is any fixed point of D, and the path of integration $\widehat{z_1 z}$ for the second integral lies in D. Then, according to Theorem 13.5, $\varphi(z)$ is a single-valued analytic function on D such that $\varphi'(z) = f(z)$, while by the argument just given, the integral

$$\int_{z_0}^{z_1} f(\zeta)\, d\zeta \qquad (\widehat{z_0 z_1} \subset G)$$

takes at least two different values, thereby generating at least two different single-valued branches of $F(z)$ on D, whose values differ by a constant. Obviously, every such branch is a primitive of $f(z)$ on D, and the proof is complete.

Remark. Actually, *every* value of the integral

$$\int_{z_0}^{z} f(\zeta)\, d\zeta$$

is of the form (13.37), and therefore belongs to one of the single-valued branches constructed in the proof of Theorem 13.7. In fact, given a curve $\Gamma \subset G$ joining z_0 to z, let $\gamma \subset D$ be any curve joining z_1 to z. Then

$$\int_{\Gamma} f(\zeta)\, d\zeta = \int_{\Gamma - \gamma} f(\zeta)\, d\zeta + \int_{\gamma} f(\zeta)\, d\zeta = \int_{\Gamma - \gamma} f(\zeta)\, d\zeta + \varphi(z),$$

where $\Gamma - \gamma \subset G$ is a curve joining z_0 to z_1.

Example. Let Π_e be the extended complex plane, and let $R_- : x < 0$, $y = 0$ be the negative real axis. For the function $f(z)$, the multiply connected domain G, the simply connected subdomain $D \subset G$ and the points z_0, z_1 figuring in Theorem 13.7 and its proof, we choose

$$f(z) = \frac{1}{z}, \quad G = \Pi_e - \{0\} - \{\infty\}, \quad D = \Pi_e - \{0\} - R_-, \quad z_0 = z_1 = 1.$$

Thus G is the "punctured plane" with the origin and point at infinity deleted, and D is the "cut plane" with the origin and all negative real numbers deleted. In this case,

$$F(z) = \int_{z_0}^{z} f(\zeta)\, d\zeta = \int_{z_0}^{z_1} f(\zeta)\, d\zeta + \int_{z_1}^{z} f(\zeta)\, d\zeta = \int_{L} \frac{d\zeta}{\zeta} + \int_{1}^{z} \frac{d\zeta}{\zeta},$$

where L is any closed rectifiable curve passing through the point $\zeta = 1$, and the integral

$$\varphi(z) = \int_{1}^{z} \frac{d\zeta}{\zeta}$$

is evaluated along any rectifiable curve joining 1 to z and lying entirely in D. According to the corollary to Theorem 13.6,

$$\varphi(z) = \Phi(z) - \Phi(1),$$

where $\Phi(z)$ is an arbitrary primitive of $f(z) = 1/z$ on D. As this primitive, we can choose the principal value of the logarithm, i.e., the single-valued branch of the logarithm which takes the value 0 at the point 1 (see Sec. 56). Then

$$\Phi(z) = \ln z = \ln |z| + i \arg z,$$

and

$$\varphi(z) = \Phi(z) - \Phi(1) = \ln z,$$

which implies

$$F(z) = \int_L \frac{d\zeta}{\zeta} + \varphi(z) = \int_L \frac{d\zeta}{\zeta} + \ln z.$$

Now let γ be the unit circle $|\zeta| = 1$ traversed once in the positive direction, and let

$$2\gamma = \gamma + \gamma, \quad 3\gamma = 2\gamma + \gamma, \ldots, \quad n\gamma = (n-1)\gamma + \gamma, \ldots,$$

so that $n\gamma$ denotes the unit circle γ traversed n times in the positive direction. Then, choosing $L = \pm n\gamma$ $(n = 1, 2, \ldots)$, we have

$$\begin{aligned}\int_L \frac{d\zeta}{\zeta} = \int_{\pm n\gamma} \frac{d\zeta}{\zeta} &= \pm \int_\gamma \frac{d\zeta}{\zeta} + \int_{\pm(n-1)\gamma} \frac{d\zeta}{\zeta} \\ &= \pm 2 \int_\gamma \frac{d\zeta}{\zeta} + \int_{\pm(n-2)\gamma} \frac{d\zeta}{\zeta} = \cdots = \pm n \int_\gamma \frac{d\zeta}{\zeta} = \pm 2n\pi i,\end{aligned}$$

since, as we already know (see Example 3, p. 252),

$$\int_\gamma \frac{d\zeta}{\zeta} = 2\pi i.$$

On the other hand, if L is any closed curve passing through the point 1 whose interior does not contain the origin, we have

$$\int_L \frac{d\zeta}{\zeta} = 0,$$

by Cauchy's integral theorem. Thus, given any integer $n = 0, \pm 1, \pm 2, \ldots$ and any point $z \in D$, we find that

$$F(z) = \int_L \frac{d\zeta}{\zeta} + \ln z = \ln z + 2n\pi i \tag{13.38}$$

for a suitably chosen contour L passing through the point 1, i.e., every single-valued branch of the function Ln z is a single-valued branch of the function $F(z)$ [on the domain D].

We could just as well have chosen the boundary of D to be any ray emanating from the origin. For example, let $I_- : x = 0, y < 0$ be the negative imaginary axis, and let $D' = \Pi_e - \{0\} - I_-$, where Π_e is the extended plane, as before. Then the same argument shows that (13.38) holds for any integer

$n = 0, \pm 1, \pm 2, \ldots$ and any point $z \in D'$. In fact, (13.38) holds for any integer $n = 0, \pm 1, \pm 2, \ldots$ and any point z in the multiply connected domain $G = \Pi_e - \{0\} - \{\infty\}$, since $G = D \cup D'$ In other words, any value of Ln z at any point $z \in G$ can be obtained by integrating the function $f(\zeta) = 1/\zeta$ along a suitable rectifiable curve joining the point 1 to the point z and encircling the origin a certain number of times (possibly not at all), either in the positive or in the negative direction.

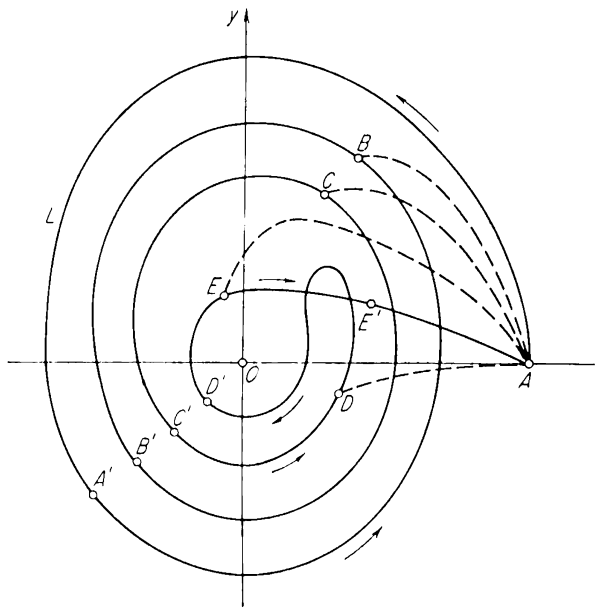

FIGURE 13.11

These considerations suggest that the multiple-valued functions $F(z)$ and Ln z are identical on G, i.e., that

$$F(z) = \int_1^z \frac{d\zeta}{\zeta} = \text{Ln } z \qquad (z \in G). \tag{13.39}$$

However, before we can write (13.39), we must still prove that *every* single-valued branch of $F(z)$ is a single-valued branch of Ln z (on some simply connected subdomain of G). Formula (13.39) will be proved if we can show that the integral

$$\int_L \frac{d\zeta}{\zeta}$$

is an integral multiple of $2\pi i$ regardless of the choice of the closed rectifiable curve L passing through the point $\zeta = 1$.[13] For example, consider the special

[13] So far, we have only considered curves L of a special kind, i.e., $L = \pm n\gamma$ or $L \subset D$, where D is a simply connected subdomain of G.

case where L is the closed curve shown in Figure 13.11. If we draw the auxiliary arcs AB, AC, AD and AE, it is clear that

$$\int_L \frac{d\zeta}{\zeta} = \int_{AA'BA} \frac{d\zeta}{\zeta} + \int_{ABB'CA} \frac{d\zeta}{\zeta} + \int_{ACC'DA} \frac{d\zeta}{\zeta} + \int_{ADD'EA} \frac{d\zeta}{\zeta} + \int_{AEE'A} \frac{d\zeta}{\zeta}$$
$$= 2\pi i + 2\pi i + 2\pi i - 2\pi i + \int_{AEE'A} \frac{d\zeta}{\zeta},$$

where we use the result of Problem 13.15. Moreover, since G contains the contour $AEE'A$ and its interior, the integral on the right vanishes, by Cauchy's integral theorem, and therefore

$$\int_L \frac{d\zeta}{\zeta} = 2 \cdot 2\pi i,$$

so that $\int_L d\zeta/\zeta$ is an integral multiple of $2\pi i$. Since this argument is of a perfectly general character, we are justified in writing (13.39), which, in a certain sense, explains why the function Ln z is multiple-valued.

PROBLEMS

13.1. Let L be a closed piecewise smooth Jordan curve enclosing an area S. Prove that

a) $\int_L x\,dz = iS$; b) $\int_L y\,dz = -S$; c) $\int_L \bar{z}\,dz = 2iS$.

Hint. Use Green's theorem.

13.2. Prove that the intersection of two or more convex sets is also convex.

13.3. Let $L = A_0A_1 \cdots A_{n-1}A_0$ be a closed polygonal Jordan curve, with interior $I(L)$. Prove that the polygon $\overline{I(L)}$ is convex if and only if none of the extensions of its sides $A_{k-1}A_k$ ($k = 1, \ldots, n$; $A_n = A_0$) intersect $I(L)$.

13.4. Prove that every convex polygon is the intersection of a finite number of (closed) half-planes.

13.5. Prove that every convex domain is simply connected.

Hint. Use Theorem 4.14.

13.6. Evaluate the integral

$$\int_{|z-a|=R} (z-a)^n \qquad (R > 0)$$

for all values of the integer n.

13.7. Suppose that $f(z)$ is analytic in the closed domain $0 \leqslant \arg z \leqslant \alpha$ ($0 < \alpha \leqslant 2\pi$) and $\lim_{z\to\infty} zf(z) = 0$. Prove that if the integral

$$J_1 = \int_0^\infty f(x)\,dx$$

exists, then so does the integral

$$J_2 = \int_L f(z)\,dz,$$

where L is the ray $z = re^{i\alpha}$, $0 \leqslant r < \infty$. Moreover, show that $J_1 = J_2$.

Hint. Cf. Problem 12.18.

13.8. Prove that

$$\int_{|z|=1} a^z\,dz = 0$$

for any single-valued branch of the function a^z.

13.9. Evaluate the integral

$$\int_{|z|=1} z^\alpha\,dz,$$

where α is an arbitrary complex number, $1^\alpha = 1$, and the path of integration begins at the point $z = 1$.

Ans. $(e^{2\alpha\pi i} - 1)/(\alpha + 1)$ if $\alpha \neq -1$; $2\pi i$ if $\alpha = -1$.

13.10. Prove that

$$\int_0^\infty \frac{x^{\alpha-1}}{1-x}\,dx = \pi \cot \alpha\pi \qquad (0 < \alpha < 1).$$

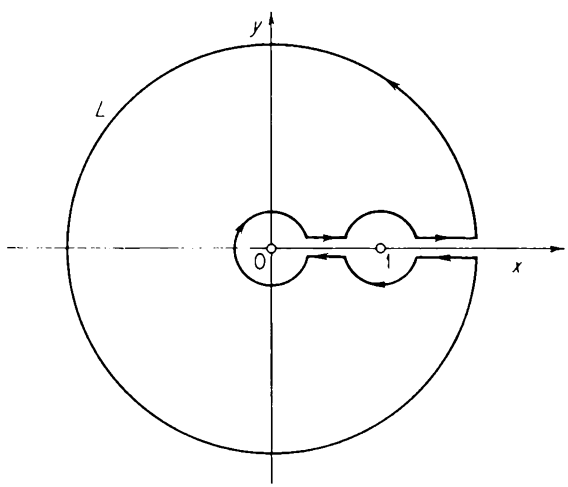

FIGURE 13.12

Hint. Define the integral as the *Cauchy principal value* (see p. 316)

$$\lim_{\varepsilon \to \infty} \left\{ \int_0^{1-\varepsilon} \frac{x^{\alpha-1}}{1-x}\,dx + \int_{1+\varepsilon}^\infty \frac{x^{\alpha-1}}{1-x}\,dx \right\}.$$

Integrate the function

$$f(z) = \frac{z^{\alpha-1}}{1-z} = \frac{e^{(\alpha-1)\ln z}}{1-z}$$

along the contour L shown in Figure 13.12, noting that $f(xe^{2\pi i}) = e^{2\alpha\pi i} f(x)$ on the lower edge of the cut along the x-axis. Also be careful about the behavior of $f(z)$ on the semicircular parts of L above and below the point $z = 1$.

13.11. Show that

$$\int_0^\infty e^{-x\cos\lambda}\, x^{\alpha-1} \cos(x \sin\lambda)\, dx = \Gamma(\alpha)\cos(\alpha\lambda),$$
$$\int_0^\infty e^{-x\sin\lambda}\, x^{\alpha-1} \sin(x \sin\lambda)\, dx = \Gamma(\alpha)\sin(\alpha\lambda),$$

where $0 < \lambda < \pi/2$, $0 < \alpha$, and

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1}\, dx$$

is the *gamma function*.[14]

Hint. Consider the integral of $z^{\alpha-1}$ along the contour L shown in Figure 13.13.

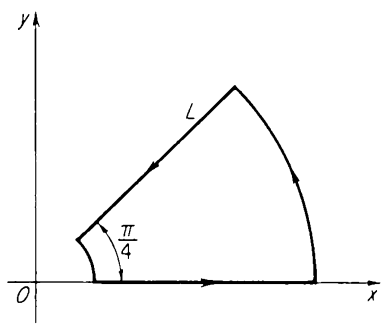

FIGURE 13.13

13.12. Prove the inequality

$$\cos\theta \geqslant 1 - \frac{2\theta}{\pi} \quad \left(0 \leqslant \theta \leqslant \frac{\pi}{2}\right).$$

13.13. Show that

$$\int_0^\infty \frac{\cos x}{x^{1-\alpha}}\, dx = \Gamma(\alpha)\cos\frac{\pi\alpha}{2},$$
$$\int_0^\infty \frac{\sin x}{x^{1-\alpha}}\, dx = \Gamma(\alpha)\sin\frac{\pi\alpha}{2},$$

where $0 < \alpha < 1$.

Hint. Use the preceding two problems.

13.14. Prove the assertion made in footnote 10. p. 280.

Hint. The new disjoint arcs need not connect the contours $\Gamma, \gamma_1, \ldots, \gamma_n$ in the same order as the old intersecting arcs.

13.15. Prove that

$$\int_L \frac{dz}{z-a} = \pm 2\pi i$$

---

[14] See e.g., D. V. Widder, *op. cit.*, Chap. 11.

if L is *any* closed rectifiable Jordan curve whose interior contains the point a, where we choose the plus sign if L is traversed in the positive direction and the minus sign if L is traversed in the negative direction.

Hint. Use Theorem 13.4.

13.16. Prove that

$$\int_L \frac{dz}{z^2 + 1} = 0$$

if L is an arbitrary closed rectifiable curve contained in the annulus $1 < |z| < R$ $(R > 1)$, but not if L is an arbitrary closed rectifiable curve contained in the domain consisting of all points z such that $z^2 + 1 \neq 0$.

Hint. Use

$$\frac{1}{z^2 + 1} = \frac{1}{2i}\left(\frac{1}{z - i} - \frac{1}{z + i}\right)$$

and the preceding problem.

13.17. Examine the behavior of the integral

$$\int_{|z-1|=R} \frac{z^4 + z^2 + 1}{z(z^2 + 1)}\, dz$$

as a function of $R > 0$.

Hint.

$$\frac{z^4 + z^2 + 1}{z(z^2 + 1)} = z + \frac{1}{z} - \frac{1}{2}\left(\frac{1}{z + i} + \frac{1}{z - i}\right).$$

13.18. Give a treatment of the function

$$F(z) = \int_0^z \frac{d\zeta}{\zeta^2 + 1}$$

paralleling that of the function

$$\operatorname{Ln} z = \int_1^z \frac{d\zeta}{\zeta},$$

given in Sec. 69. With what multiple-valued function does $F(z)$ coincide?

13.19. State and prove the analogue for complex functions of the familiar rule for partial integration of real functions.

14

CAUCHY'S INTEGRAL AND RELATED TOPICS

70. Cauchy's Integral Formula

We now prove a result of fundamental importance in the theory of analytic functions:

THEOREM 14.1 (*Cauchy's integral formula*). *If $f(z)$ is a single-valued analytic function on a domain G, and if G contains a closed rectifiable Jordan curve L and its interior $I(L)$, then*

$$\frac{1}{2\pi i}\int_L \frac{f(\zeta)}{\zeta - z}\,d\zeta = f(z) \qquad \text{if} \quad z \in I(L), \tag{14.1}$$

where the integral on the left is called Cauchy's integral.[1] *On the other hand,*

$$\frac{1}{2\pi i}\int_L \frac{f(\zeta)}{\zeta - z}\,d\zeta = 0 \qquad \text{if} \quad z \in E(L), \tag{14.2}$$

where $E(L)$ denotes the exterior of L.

Proof. First let z be an arbitrary point of the domain $I(L)$. Then the function

$$g(\zeta) = \frac{f(\zeta)}{\zeta - z},$$

[1] It is assumed that L is traversed in the direction such that $I(L)$ lies to the left of an observer moving along L.

regarded as a function of the variable point ζ, is defined and analytic on the domain $G' = G - \{z\}$, since $g(\zeta)$ is the ratio of two analytic functions on G'. Let ρ be so small that $I(L)$ contains $\overline{I(\gamma_\rho)}$, where γ_ρ is the circle with center z and radius ρ (see Figure 14.1). Then Theorem 13.4, applied to the contours L and γ_ρ, implies

$$\int_L g(\zeta)\,d\zeta = \int_{\gamma_\rho} g(\zeta)\,d\zeta$$

or

$$\int_L \frac{f(\zeta)}{\zeta - z}\,d\zeta = \int_{\gamma_\rho} \frac{f(\zeta)}{\zeta - z}\,d\zeta. \tag{14.3}$$

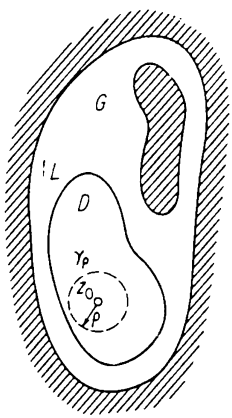

FIGURE 14.1

According to (14.3) and the construction of γ_ρ, the value of the integral

$$\int_{\gamma_\rho} \frac{f(\zeta)}{\zeta - z}\,d\zeta$$

does not change if we decrease the radius ρ, and hence

$$\int_{\gamma_\rho} \frac{f(\zeta)}{\zeta - z}\,d\zeta = \lim_{\rho \to 0} \int_{\gamma_\rho} \frac{f(\zeta)}{\zeta - z}\,d\zeta.$$

Therefore, to prove (14.1), we need only show that

$$\lim_{\rho \to 0} \int_{\gamma_\rho} \frac{f(\zeta)}{\zeta - z}\,d\zeta = 2\pi i f(z),$$

i.e., that given any $\varepsilon > 0$, there exists a $\delta = \delta(\varepsilon) > 0$ such that

$$\left| \int_{\gamma_\rho} \frac{f(\zeta)}{\zeta - z}\,d\zeta - 2\pi i f(z) \right| < \varepsilon \tag{14.4}$$

whenever $\rho < \delta$. Since

$$\int_{\gamma_\rho} \frac{d\zeta}{\zeta - z} = 2\pi i$$

(see Example 3, p. 252), we can write the left-hand side of (14.4) in the form

$$\begin{aligned} \left| \int_{\gamma_\rho} \frac{f(\zeta)}{\zeta - z}\,d\zeta - 2\pi i f(z) \right| &= \left| \int_{\gamma_\rho} \frac{f(\zeta)}{\zeta - z}\,d\zeta - f(z) \int_{\gamma_\rho} \frac{d\zeta}{\zeta - z} \right| \\ &= \left| \int_{\gamma_\rho} \frac{f(\zeta) - f(z)}{\zeta - z}\,d\zeta \right|. \end{aligned}$$

Because of the continuity of $f(z)$, given any $\varepsilon > 0$, we can find a $\delta = \delta(\varepsilon) > 0$ such that

$$|f(\zeta) - f(z)| < \frac{\varepsilon}{2\pi}$$

whenever

$$|\zeta - z| = \rho < \delta.$$

Therefore

$$\left| \int_{\gamma_\rho} \frac{f(\zeta)}{\zeta - z}\, d\zeta - 2\pi i f(z) \right| < \frac{\varepsilon/2\pi}{\rho}\, 2\pi\rho = \varepsilon$$

whenever $\rho < \delta$. This proves (14.4) and hence (14.1).

To prove (14.2), we note that

$$\int_L \frac{f(\zeta)}{\zeta - z}\, d\zeta = 0$$

if $z \in E(L)$. This follows from Theorem 13.1″, p. 269 and the fact that

$$g(\zeta) = \frac{f(\zeta)}{\zeta - z},$$

regarded as a function of ζ, is analytic on the domain $G' = G - \{z\}$, which now contains both L and $I(L)$.

71. Some Consequences of Theorem 14.1

THEOREM 14.2. *If $f(z)$ is an analytic function on a domain G, and if G contains the circle $\gamma_\rho: |z - z_0| = \rho$ and its interior $I(\gamma_\rho)$, then*

$$f(z_0) = \frac{1}{2\pi} \int_0^{2\pi} f(z_0 + \rho e^{i\theta})\, d\theta, \tag{14.5}$$

i.e., the value of $f(z)$ at the point z_0 equals the average of its values on the circle γ_ρ (with center z_0).[2]

Proof. The equation of γ_ρ is

$$z = z_0 + \rho e^{i\theta} \qquad (0 \leqslant \theta \leqslant 2\pi). \tag{14.6}$$

It follows from (14.6) and Cauchy's integral formula (14.1) that

$$f(z_0) = \frac{1}{2\pi i} \int_0^{2\pi} \frac{f(z_0 + \rho e^{i\theta}) i\rho e^{i\theta}}{\rho e^{i\theta}}\, d\theta = \frac{1}{2\pi} \int_0^{2\pi} f(z_0 + \rho e^{i\theta})\, d\theta,$$

which proves (14.5).

COROLLARY. *If $f(z)$ is an analytic function on a domain G, and if G contains the circle $\gamma_\rho: |z - z_0| = \rho$ and its interior $I(\gamma_\rho)$, then*

$$|f(z_0)| \leqslant M(\rho) \equiv \max_{z \in \gamma_\rho} |f(z)|. \tag{14.7}$$

---

[2] From now on, we assume that all functions are single-valued, unless the contrary is explicitly stated.

Proof. According to (14.5) and the definition of $M(\rho)$,

$$|f(z_0)| = \left|\frac{1}{2\pi}\int_0^{2\pi} f(z_0 + \rho e^{i\theta})\, d\theta\right| \leqslant \frac{1}{2\pi}\int_0^{2\pi} |f(z_0 + \rho e^{i\theta})|\, d0 \leqslant M(\rho),$$

as asserted.

THEOREM 14.3. *If $f(z)$ is an analytic function on a domain G, then $|f(z)|$ cannot have a relative maximum in the strict sense at any point of G.*[3]

Proof. Let γ_ρ be the same as in the corollary. Since $|f(z)|$ is continuous on γ_ρ, there is a point $\zeta \in \gamma_\rho$ such that $|f(z_0)| \leqslant M(\rho) = |f(\zeta)|$ [see Theorem 3.16]. Therefore, since the radius of γ_ρ can be chosen arbitrarily small, in any neighborhood of the point $z_0 \in G$ we can always find another point where the inequality $|f(z_0)| > |f(z)|$ does not hold.

Remark. In Sec. 83 we shall strengthen Theorem 14.3 by showing that the word "strict" can be replaced by the word "wide," provided that $f(z) \not\equiv$ const on G, thereby proving the *maximum modulus principle*. For the time being, we shall prove some related but slightly weaker results.

THEOREM 14.4. *If $f(z) \not\equiv$ const is an analytic function on a domain G, then*

$$|f(z)| < M = \sup_{z' \in G} |f(z')| \qquad (z \in G), \tag{14.8}$$

i.e., $|f(z)|$ cannot achieve its least upper bound on G.

Proof. We can assume that $0 < M < \infty$, since $f(z) \equiv 0$ if $M = 0$ and the theorem is obvious if $M = \infty$. Suppose that contrary to hypothesis, $|f(z_0)| = M$ for some $z_0 \in G$, and let ρ be so small that G contains the circle $\gamma_\rho: |z - z_0| = \rho$ and its interior $I(\gamma_\rho)$. Then, according to (14.5), we have

$$|f(z_0)| = M = \left|\frac{1}{2\pi}\int_0^{2\pi} f(z_0 + \rho e^{i\theta})\, d\theta\right| \leqslant \frac{1}{2\pi}\int_0^{2\pi} |f(z_0 + \rho e^{i\theta})|\, d\theta. \tag{14.9}$$

Obviously $|f(z_0 + \rho e^{i\theta})| \leqslant M$, $0 \leqslant \theta \leqslant 2\pi$, but the strict inequality cannot hold for any θ. In fact, if

$$\mu = |f(z_0 + \rho e^{i\theta_0})| < M,$$

then, because of the continuity of $|f(z)|$ on γ_ρ, the interval $[0, 2\pi]$ contains a subinterval α of length $\delta > 0$ such that

$$|f(z_0 + \rho e^{i\theta})| < \frac{M + \mu}{2} \qquad \text{if} \quad \theta \in \alpha.$$

---

[3] A real-valued function $g(z)$ defined on a set E is said to have a *relative maximum (in the wide sense)* at a point $z_0 \in E$ if there is a neighborhood $\mathcal{N}(z_0)$ such that $g(z_0) \geqslant g(z)$ for all $z \in \mathcal{N}(z_0) \cap E$. If the sign $\geqslant$ can be replaced by $>$, then $g(z)$ is said to have a relative maximum *in the strict sense* at z_0.

But then (14.9) implies

$$M \leqslant \frac{1}{2\pi}\int_0^{2\pi} |f(z_0 + \rho e^{i\theta})|\, d\theta = \frac{1}{2\pi}\int_{[0,\,2\pi]-\alpha} |f(z_0 + \rho e^{i\theta})|\, d\theta + \frac{1}{2\pi}\int_\alpha |f(z_0 + \rho e^{i\theta})|\, d\theta \leqslant \frac{1}{2\pi} M(2\pi - \delta) + \frac{1}{2\pi}\frac{M + \mu}{2}\delta < M,$$

which is impossible. Therefore, if $|f(z_0)| = M$,

$$|f(z_0 + \rho e^{i\theta})| = M \qquad (0 \leqslant \theta \leqslant 2\pi)$$

for all sufficiently small ρ.

Now let E be the set of all $z \in G$ such that $|f(z)| = M$. Clearly E is open, and moreover any limit point of E belonging to G also belongs to E, because of the continuity of $|f(z)|$. Therefore $E = G$, by Theorem 4.16. In other words, if

$$|f(z_0)| = M = \sup_{z \in G} |f(z)|$$

at a single point $z_0 \in G$, it follows that $|f(z)| \equiv |u(x, y) + iv(x, y)| = M$ or

$$|f(z)|^2 = [u(x, y)]^2 + [v(x, y)]^2 = M^2 \tag{14.10}$$

for *all* $z \in G$. Differentiating (14.10) with respect to x and y, we obtain

$$u\frac{\partial u}{\partial x} + v\frac{\partial v}{\partial x} = 0, \qquad u\frac{\partial u}{\partial y} + v\frac{\partial v}{\partial y} = 0,$$

and since u and v cannot vanish simultaneously ($M \neq 0$),

$$\frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial v}{\partial x}\frac{\partial u}{\partial y} = 0. \tag{14.11}$$

Using the Cauchy-Riemann equations (see Theorem 7.2), we can write (14.11) in the form

$$\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 = 0,$$

which implies

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = 0$$

and hence

$$\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} = 0$$

(everywhere in G), where we have again used the Cauchy-Riemann equations. But then

$$u(x, y) \equiv C_1, \quad v(x, y) \equiv C_2, \quad f(z) \equiv C_1 + iC_2 \equiv C,$$

where C_1, C_2 and C are constants. This completes the proof of the theorem.

COROLLARY 1. *Let G be a domain, and let Γ be the boundary of G. Suppose $f(z)$ is analytic on G and continuous on $\bar{G} = G \cup \Gamma$. Then there is a point $\zeta \in \Gamma$ such that*

$$|f(\zeta)| = M_0 = \sup_{z \in G} |f(z)|.$$

Proof. Since $|f(z)|$ is continuous on $\bar{G}$, there is a point $\zeta \in \bar{G}$ such that $|f(\zeta)| = M_0$ (see Theorem 3.16). But according to (14.8), $|f(z)| < M_0$ on G itself, since

$$\sup_{z \in G} |f(z)| \leqslant M_0.$$

Therefore $\zeta \in \bar{G} - G$, which means that $\zeta \in \Gamma$, as asserted.

COROLLARY 2. *If $f(z)$ is an analytic function on a domain G, and if G contains the circle $\gamma_\rho: |z - z_0| = \rho$ and its interior $I(\gamma_\rho)$, then*[4]

$$|f(z_0)| < M(\rho) = \max_{z \in \gamma_\rho} |f(z)|, \tag{14.12}$$

provided that $f(z) \not\equiv$ const on $I(\gamma_\rho)$. Moreover, if $\rho' < \rho$ and if $\gamma_{\rho'}$ is the circle $|z - z_0| = \rho'$, then

$$\max_{z \in \gamma_{\rho'}} |f(z)| = M(\rho') < M(\rho),$$

i.e., $M(\rho)$ is a strictly increasing function of ρ.[5]

Proof. First we note that

$$|f(z_0)| < \sup_{|z - z_0| < \rho} |f(z)| \leqslant \sup_{|z - z_0| \leqslant \rho} |f(z)| = M(\rho),$$

where we have used Corollary 1. Next we observe that if $\rho' < \rho$, then $M(\rho') = |f(\zeta)|$, where $\zeta \in \gamma_{\rho'}$. It follows that

$$M(\rho') = |f(\zeta)| < \sup_{|z - z_0| < \rho} |f(z)| \leqslant M(\rho).$$

Remark. Because of (14.12), the same argument as used to derive Theorem 14.3 would immediately imply the maximum modulus principle (see the remark on p. 296), if it were only known that $f(z) \not\equiv$ const on a domain G implies $f(z) \not\equiv$ const on every open disk (neighborhood) contained in G. In Sec. 83 we will prove that this is indeed the case.

---

[4] This result should be compared with the corollary to Theorem 14.2. In particular, note the difference between (14.12) and (14.7).

[5] Cf. Problem 4.7. Obviously, if $z_0 \in G$, then

$$M(\rho) = \max_{z \in \gamma_\rho} |f(z)|$$

is defined for all ρ such that $0 < \rho < \Delta$, where Δ is the distance between z_0 and the boundary of G.

72. Integrals of the Cauchy Type. Cauchy's Inequalities

We now introduce an important generalization of Cauchy's integral: By an *integral of the Cauchy type* is meant an expression of the form

$$\frac{1}{2\pi i}\int_L \frac{\varphi(\zeta)}{\zeta - z}\,d\zeta \qquad (z \notin L), \tag{14.13}$$

where L is a rectifiable curve (not necessarily closed), and $\varphi(z)$ is a function, often called the *density*, which is continuous on L.[6] Obviously, (14.13) reduces to Cauchy's integral if L is a closed rectifiable Jordan curve and if $\varphi(z)$ is a function which is analytic on a domain containing L and its interior. It is an immediate consequence of (14.13) that an integral of the Cauchy type defines a single-valued function $f(z)$ on every domain which contains no points of L.

Example. Let α be a closed interval of the real axis, and let $\varphi(x) \not\equiv 0$ be a continuous function of a real variable defined on α. Then the integral

$$\frac{1}{2\pi i}\int_\alpha \frac{\varphi(x)}{x - z}\,dx \qquad (\operatorname{Im} z \neq 0),$$

and the special case

$$\frac{1}{2\pi i}\int_\alpha \frac{dx}{x - z} \qquad (\operatorname{Im} z \neq 0),$$

are integrals of the Cauchy type.

THEOREM 14.5. *Every integral of the Cauchy type*

$$f(z) = \frac{1}{2\pi i}\int_L \frac{\varphi(\zeta)}{\zeta - z}\,d\zeta \tag{14.14}$$

defines an infinitely differentiable function $f(z)$ on any domain G containing no points of L. Moreover, the n'th derivative of $f(z)$ is obtained by differentiating the integrand of (14.14) *n times with respect to z, i.e.,*

$$f^{(n)}(z) = \frac{n!}{2\pi i}\int_L \frac{\varphi(\zeta)}{(\zeta - z)^{n+1}}\,d\zeta \qquad (n = 1, 2, \ldots). \tag{14.15}$$

Proof. We will prove (14.15) by mathematical induction. First we note that (14.15) holds for $n = 0$, since $f^{(0)}(z) = f(z)$ by definition (recall that $0! = 1$). Next, assuming that (14.15) holds for some non-negative integer n, we show that (14.15) also holds for $n + 1$. This will be accomplished by direct calculation of the quantity

$$f^{(n+1)}(z_0) = \lim_{z \to z_0} \frac{f^{(n)}(z) - f^{(n)}(z_0)}{z - z_0},$$

---

[6] There is no contradiction between writing $z \notin L$ in (14.13) and using $\varphi(z)$ to denote a function defined on L. A fussier approach would be to write just φ instead of $\varphi(z)$, thereby distinguishing between a function and its value at a point (see p. 24).

where z_0 is an arbitrary point of the domain G. For sufficiently small ρ, G contains the circle $\gamma_\rho: |z - z_0| = \rho$ and its interior $I(\gamma_\rho)$. Let $\delta > 0$ be the distance between γ_ρ and the curve L. Moreover, let $\Gamma: |z| = R$ be a circle (with center at the origin) such that $\overline{I(\gamma_\rho)} \subset I(\Gamma)$, $L \subset I(\Gamma)$. Then

$$f^{(n)}(z) - f^{(n)}(z_0) = \frac{n!}{2\pi i}\int_L \varphi(\zeta)\frac{(\zeta - z_0)^{n+1} - (\zeta - z)^{n+1}}{(\zeta - z_0)^{n+1}(\zeta - z)^{n+1}}\,d\zeta \quad (14.16)$$

for any $z \in \overline{I(\gamma_\rho)}$. Writing $\zeta - z_0 = t$, $z - z_0 = h$, so that $\zeta - z = t - h$, and using (14.16) and the algebraic identity

$$a^{n+1} - b^{n+1} = (a - b)(a^n + a^{n-1}b + \cdots + b^n),$$

we find that

$$\frac{f^{(n)}(z_0 + h) - f^{(n)}(z_0)}{h} = \frac{n!}{2\pi i}\int_L \varphi(\zeta)\frac{(t - h)^n + t(t - h)^{n-1} + \cdots + t^n}{t^{n+1}(t - h)^{n+1}}\,d\zeta, \quad (14.17)$$

Since our objective is to show that (14.17) approaches the limit

$$\psi(z_0) = \frac{(n + 1)!}{2\pi i}\int_L \frac{\varphi(\zeta)}{(\zeta - z_0)^{n+2}}\,d\zeta = \frac{(n + 1)!}{2\pi i}\int_L \frac{\varphi(\zeta)}{t^{n+2}}\,d\zeta$$

as $h \to 0$, we now examine the difference

$$\begin{aligned}\frac{f^{(n)}(z_0 + h) - f^{(n)}(z_0)}{h} - \psi(z_0) &= \frac{n!}{2\pi i}\int_L \varphi(\zeta)\frac{t(t - h)^n + t^2(t - h)^{n-1} + \cdots + t^{n+1} - (n + 1)(t - h)^{n+1}}{t^{n+2}(t - h)^{n+1}}\,d\zeta \\ &= \frac{n!h}{2\pi i}\int_L \varphi(\zeta)\frac{(t-h)^n + [t + (t-h)](t-h)^{n-1} + \cdots + [t^n + t^{n-1}(t-h) + \cdots + (t-h)^n]}{t^{n+2}(t - h)^{n+1}}\,d\zeta\end{aligned} \quad (14.18)$$

[note that $h = t - (t - h)$]. By construction, we have the estimates

$$\delta < |t| = |\zeta - z_0| < 2R, \qquad \delta \leqslant |t - h| = |\zeta - z| < 2R,$$

(draw a figure). Therefore it follows from (14.18) that

$$\left|\frac{f^{(n)}(z_0 + h) - f^{(n)}(z_0)}{h} - \psi(z_0)\right| \leqslant \frac{n!\,|h|M}{2\pi}\,\frac{(2R)^n + 2(2R)^n + 3(2R)^n + \cdots + (n + 1)(2R)^n}{\delta^{2n+3}}\,l, \quad (14.19)$$

where l is the length of L and

$$M = \max_{\zeta \in L} |\varphi(\zeta)|.$$

But the right-hand side of (14.19) obviously approaches zero as $h \to 0$. In other words,

$$f^{(n+1)}(z_0) = \lim_{h \to 0} \frac{f^{(n)}(z_0 + h) - f^{(n)}(z_0)}{h}$$
$$= \psi(z_0) = \frac{(n+1)!}{2\pi} \int_L \frac{\varphi(\zeta)}{(\zeta - z_0)^{n+2}}\, d\zeta,$$

which completes the proof.

COROLLARY 1. *If $f(z)$ is an analytic function on a domain G, then $f(z)$ is infinitely differentiable on G.*

Proof. Given any point $z_0 \in G$, we can choose $\rho > 0$ such that G contains the circle $\gamma_\rho: |z - z_0| = \rho$ and its interior $I(\gamma_\rho)$. Therefore, according to Theorem 14.1,

$$f(z_0) = \frac{1}{2\pi i} \int_{\gamma_\rho} \frac{f(\zeta)}{\zeta - z_0}\, d\zeta,$$

i.e., $f(z)$ is represented by an integral of the Cauchy type (in fact, by Cauchy's integral) at the point z_0. The proof now follows from Theorem 14.5 and the fact that the point $z_0 \in G$ is arbitrary. In particular, we note that

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \int_{\gamma_\rho} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\, d\zeta \qquad (n = 1, 2, \ldots), \tag{14.20}$$

where γ_ρ can be replaced by any other closed rectifiable Jordan curve L such that $z_0 \in I(L)$ and $\overline{I(L)} \subset G$.

COROLLARY 2. *If $f(z)$ is an analytic function on a domain G, then every derivative*

$$f^{(n)}(z) = \frac{d^n f(z)}{dz^n} \qquad (n = 1, 2, \ldots)$$

is analytic on G.

The next result serves as the converse of Cauchy's integral theorem:

THEOREM 14.6 (*Morera's theorem*). *Let $f(z)$ be a continuous function on a simply connected domain G, and suppose that*

$$\int_L f(z)\, dz = 0, \tag{14.21}$$

where L is any closed rectifiable curve contained in G. Then $f(z)$ is analytic on G.

Proof. If z_0 and z are any two points in G, it follows from (14.21) and the argument given at the beginning of Sec. 68 that the integral

$$F(z) = \int_{z_0}^{z} f(\zeta)\, d\zeta$$

is path-independent. Therefore, according to Theorem 13.5,[7] $F(z)$ is analytic on G, with derivative $F'(z) = f(z)$. But then Theorem 14.5, Corollary 2 implies that $f(z)$ itself is analytic on G.

Remark. In the statement of Theorem 14.6 we need only assume that (14.21) vanishes for an arbitrary triangular contour $L \subset G$. This follows by the same argument used to prove Theorem 13.1 (cf. Sec. 65, Steps 2 through 6).

THEOREM 14.7. *Let $f(z)$ be an analytic function on a domain G, and suppose G contains the circle $\gamma_\rho: |z - z_0| = \rho$ and its interior $I(\gamma_\rho)$. Then*

$$|f^{(n)}(z_0)| \leqslant n! \frac{M(\rho)}{\rho^n} \qquad (n = 0, 1, 2, \ldots), \tag{14.22}$$

where

$$M(\rho) = \max_{z \in \gamma_\rho} |f(z)|. \tag{14.23}$$

Proof. The inequalities (14.22), known as the *Cauchy inequalities*, are an immediate consequence of the representation (14.20), which implies

$$|f^{(n)}(z_0)| \leqslant \frac{n!}{2\pi} \frac{M(\rho)}{\rho^{n+1}} 2\pi\rho = n! \frac{M(\rho)}{\rho^n}.$$

The fact that (14.22) holds for $n = 0$ has already been proved (see Theorem 14.2, Corollary). It should be noted that knowledge of the single function (14.23) is sufficient to write estimates for *all* the derivatives of $f(z)$.

Remark. For given n and z_0, the estimate (14.22) depends on the quantity ρ, which can vary between 0 and Δ, where Δ is the distance between z_0 and the boundary of G. As ρ increases, both the numerator and denominator of the ratio $M(\rho)/\rho^n$ increase.[8] The estimate (14.22) can be improved by taking ρ to be the value for which the ratio $M(\rho)/\rho^n$ achieves its minimum in the interval $(0, \Delta)$.

Example. Let G be the interior of the unit circle $|z| = 1$, and suppose that

$$M(\rho) = \frac{1}{1 - \rho}. \tag{14.24}$$

---

[7] Note that in proving Theorem 13.5, the analyticity of $f(z)$ is only used to establish (14.21), which in the present case is *assumed.*

[8] Cf. Theorem 14.4, Corollary 2.

Then, according to (14.22),

$$|f^{(n)}(0)| \leqslant n! \frac{1}{(1-\rho)\rho^n} \qquad (0 < \rho < 1). \tag{14.25}$$

To minimize the right-hand side of (14.25), we maximize $(1-\rho)\rho^n$ in the interval (0, 1). The maximum in question occurs for

$$\rho = \frac{n}{n+1},$$

and then

$$\frac{1}{(1-\rho)\rho^n} = (n+1)\left(1+\frac{1}{n}\right)^n < e(n+1).$$

It follows that

$$|f^{(n)}(0)| \leqslant e(n+1)! \tag{14.26}$$

In the special case of the function

$$f(z) = \frac{1}{1-z},$$

which is analytic on the interior of the unit circle, (14.24) is satisfied, and in fact,

$$f^{(n)}(z) = \frac{n!}{(1-z)^{n+1}}, \quad f^{(n)}(0) = n!,$$

which is consistent with the estimate (14.26).

THEOREM 14.8. *Let $f(z)$ be an analytic function on a domain G, and let z_0 be an arbitrary point of G. Then*[9]

$$\overline{\lim_{n\to\infty}} \sqrt[n]{\frac{|f^{(n)}(z_0)|}{n!}} \leqslant \frac{1}{\Delta}, \tag{14.27}$$

where Δ is the distance between z_0 and the boundary of G.

Proof. To establish the estimate (14.27), known as the *Cauchy-Hadamard inequality*, we write (14.22) in the form

$$\sqrt[n]{\frac{|f^{(n)}(z_0)|}{n!}} \leqslant \frac{\sqrt[n]{M(\rho)}}{\rho}, \tag{14.28}$$

---

[9] If $\{a_n\}$ is a bounded sequence of real numbers, then, according to Theorem 3.1, $\{a_n\}$ has at least one limit point. The largest of these limit points is called the *limit superior* or *upper limit* of the sequence, denoted by

$$\overline{\lim_{n\to\infty}} a_n \quad \text{or} \quad \limsup_{n\to\infty} a_n$$

(one often omits writing $n \to \infty$). Obviously,

$$\overline{\lim_{n\to\infty}} a_n = \lim_{n\to\infty} a_n$$

in the case where $\{a_n\}$ has a limit.

where ρ is any positive number less than Δ. Since

$$\lim_{n \to \infty} \sqrt[n]{M(\rho)} = 1,$$

(14.28) implies

$$\overline{\lim_{n \to \infty}} \sqrt[n]{\frac{|f^{(n)}(z_0)|}{n!}} \leqslant \overline{\lim_{n \to \infty}} \frac{\sqrt[n]{M(\rho)}}{\rho} = \frac{1}{\rho}, \tag{14.29}$$

where $0 < \rho < \Delta$. Letting $\rho \to \Delta$ in (14.29), we obtain (14.27).

Remark. The left-hand side of (14.27), which depends on the values of the derivatives of $f(z)$ at the point z_0, cannot be large if the distance Δ from z_0 to the boundary of G is large. In particular, if $f(z)$ is an entire function, G is the whole plane with the exception of the point at infinity, and then $\Delta = \infty$ for every point $z_0 \in G$. Therefore, for entire functions we have

$$\lim_{n \to \infty} \sqrt[n]{\frac{|f^{(n)}(z_0)|}{n!}} = 0$$

for every finite z_0.

Example. Consider the entire function

$$f(z) = e^z.$$

In this case,

$$f^{(n)}(z) = e^z$$

for all n, and hence

$$|f^{(n)}(z)| = |e^z| = e^x.$$

Moreover, if $k = [n/2]$,[10] then

$$n! \geqslant n(n-1) \cdots (n-k) > k^{n-k}$$

and

$$\sqrt[n]{n!} > k^{1-(k/n)} \geqslant \sqrt{k}.$$

It follows that

$$\sqrt[n]{\frac{|f^{(n)}(z_0)|}{n!}} < \frac{e^{x/n}}{\sqrt{[n/2]}} \to 0$$

as $n \to \infty$, in keeping with the above remark.

Finally, we derive the rule for making a change of variable in an integral of a complex function:

THEOREM 14.9. *Let $f(z)$ be an analytic function on a domain G, and let L be a rectifiable curve contained in G, with image $\Lambda = f(L)$ under the mapping $w = f(z)$. Then Λ is rectifiable and*

$$\int_\Lambda \Phi(w)\, dw = \int_L \Phi[f(z)] f'(z)\, dz, \tag{14.30}$$

where $\Phi(w)$ is any continuous function on Λ.

---

[10] By $[\nu]$ is meant the *integral part* of ν, i.e., the largest integer $\leqslant \nu$.

Proof. Since $f(z)$ is analytic on G, $f'(z)$ is analytic and hence continuous on G, so that (14.30) is meaningful, provided that Λ is rectifiable. If L has the equation

$$z = \lambda(t) \qquad (a \leqslant t \leqslant b),$$

then Λ has the equation

$$w = f[\lambda(t)] \qquad (a \leqslant t \leqslant b).$$

As in Sec. 60, let $\mathscr{P} = \{t_0, t_1, \ldots, t_n\}$ be any partition of the interval $[a, b]$, where

$$a = t_0 < t_1 < \cdots < t_n = b,$$

and let

$$z_k = \lambda(t_k), \quad w_k = f[\lambda(t_k)] \qquad (k = 0, 1, \ldots, n).$$

Then, taking the least upper bound with respect to all possible partitions $\mathscr{P}$, we find that

$$\sup_{\mathscr{P}} \sum_{k=1}^{n} |w_k - w_{k-1}| = \sup_{\mathscr{P}} \sum_{k=1}^{n} \left| \int_{\sigma_k} f'(z)\, dz \right| \leqslant M \sum_{k=1}^{n} l_k = Ml,$$

where σ_k $(k = 1, \ldots, n)$ is the arc of the curve L joining the point z_{k-1} to the point z_k, l_k is the length of σ_k, l is the length of L, and

$$M = \max_{z \in L} |f'(z)|.$$

Therefore

$$\sup_{\mathscr{P}} \sum_{k=1}^{n} |w_k - w_{k-1}| < \infty,$$

so that Λ is rectifiable, as asserted.

To prove (14.30), we note that

$$\begin{aligned} \int_{\Lambda} \Phi(w)\, dw &= \lim_{|\mathscr{P}| \to 0} \sum_{k=1}^{n} \Phi(w_k)(w_k - w_{k-1}) \\ &= \lim_{|\mathscr{P}| \to 0} \sum_{k=1}^{n} \Phi[f(z_k)] \int_{\sigma_k} f'(z)\, dz, \end{aligned}$$

while

$$\int_{L} \Phi[f(z)] f'(z)\, dz = \lim_{|\mathscr{P}| \to 0} \sum_{k=1}^{n} \int_{\sigma_k} \Phi[f(z)] f'(z)\, dz,$$

and consequently

$$\begin{aligned} &\int_{\Lambda} \Phi(w)\, dw - \int_{L} \Phi[f(z)] f'(z)\, dz \\ &\qquad = \lim_{|\mathscr{P}| \to 0} \sum_{k=1}^{n} \int_{\sigma_k} \{\Phi[f(z_k)] - \Phi[f(z)]\} f'(z)\, dz. \end{aligned} \tag{14.31}$$

Here, as on p. 245, $|\mathscr{P}|$ denotes the largest of the numbers

$$t_1 - t_0, t_2 - t_1, \ldots, t_n - t_{n-1},$$

i.e., the norm of the partition $\mathscr{P}$. Since $\Phi(w)$ is continuous on Λ, $\Phi[f(z)]$ is continuous on L, and hence, given any $\varepsilon > 0$, there exists a number $\delta = \delta(\varepsilon) > 0$ such that if $|\mathscr{P}| < \delta$, then

$$\max_{z \in \sigma_k} |\Phi[f(z_k)] - \Phi[f(z)]| < \varepsilon$$

for all $k = 1, \ldots, n$. It follows that

$$\sum_{k=1}^{n} \int_{\sigma_k} \{\Phi[f(z_k)] - \Phi[f(z)]\} f'(z)\, dz \leqslant M\varepsilon \sum_{k=1}^{n} l_k = M\varepsilon l$$

if $|\mathscr{P}| < \delta$. Therefore, since ε is arbitrary, the limit in the right-hand side of (14.31) vanishes, and the theorem is proved.

*73. Boundary Values of Integrals of the Cauchy Type

In this section and the next, we shall study the limiting behavior of integrals of the Cauchy type, i.e., integrals of the form

$$f(z) = \frac{1}{2\pi i} \int_L \frac{\varphi(\zeta)}{\zeta - z}\, d\zeta \qquad (z \in E,\ z \notin L), \tag{14.32}$$

with various assumptions concerning the curve L, the density $\varphi(z)$ and the set E, as the point $z \in E$ approaches a point $z_0 \in L$. Such a limit

$$\lim_{\substack{z \to z_0 \\ z \in E}} f(z)$$

will be called a *boundary value* of $f(z)$. Our first result is an immediate consequence of Theorem 14.1:

THEOREM 14.10. *Let L be a closed rectifiable curve, let $\varphi(z)$ be analytic on a domain G containing L and its interior $I(L)$, and let z_0 be a point of L. Then*

$$\lim_{\substack{z \to z_0 \\ z \in I(L)}} f(z) = \varphi(z_0), \tag{14.33}$$

while

$$\lim_{\substack{z \to z_0 \\ z \in E(L)}} f(z) = 0. \tag{14.34}$$

Remark. Recall that it is assumed that an observer moving along L finds $I(L)$ on his left and $E(L)$ [the exterior of L] on his right. This suggests calling (14.33) the *left-hand* boundary value of $f(z)$, denoted by $f_l(z_0)$, and (14.34) the *right-hand* boundary value of $f(z)$, denoted by $f_r(z_0)$. Then, combining (14.33) and (14.34), we find that

$$f_l(z_0) - f_r(z_0) = \varphi(z_0). \tag{14.35}$$

THEOREM 14.11. *Let L be a rectifiable Jordan curve (not necessarily closed), let $\varphi(z)$ be analytic at every point of L, and let z_0 be a point of L.*[11] *Then formula* (14.35) *remains valid with a suitable definition of the boundary values $f_l(z_0)$ and $f_r(z_0)$.*

Proof. Suppose the equation of L is

$$z = \lambda(t) \qquad (a \leqslant t \leqslant b),$$

and let $z_0 = \lambda(t_0)$, where $t_0 \neq a$, $t_0 \neq b$, if $a \neq b$. Then there exists a circle γ with center z_0 such that $\varphi(z)$ is analytic on $\overline{I(\gamma)}$ and such that γ contains at least one point $z = \lambda(t)$ with $a \leqslant t < t_0$ and at least one point with $t_0 < t \leqslant b$. Let t_1 be the largest value of t satisfying both $a \leqslant t < t_0$ and $\lambda(t) \in \gamma$, and let t_2 be the smallest value of t satisfying both $t_0 \leqslant t < b$ and $\lambda(t) \in \gamma$. In other words, if L is traversed in the positive direction (i.e., the direction of increasing t), $z_1 = \lambda(t_1)$ is the "last point" of γ encountered before reaching z_0, and $z_2 = \lambda(t_2)$ is the "first point" of γ encountered after leaving z_0. Let $\sigma = \overset{\frown}{z_1 z_2}$ be the arc of the curve L joining the point $z_1 = \lambda(t_1)$ to the point $z_2 = \lambda(t_2)$. Clearly, $z_0 \in \sigma$ and $\sigma \subset \overline{I(\gamma)}$. The points z_1 and z_2 divide the circle γ into two arcs γ_l and γ_r joining z_1 to z_2, as shown in Figure 14.2. Together with σ, these arcs form two closed rectifiable Jordan curves $\Gamma_l = \sigma - \gamma_l$ and $\Gamma_r = \sigma - \gamma_r$ (cf. footnote 11, p. 280) such that $I(\Gamma_l) \subset I(\gamma)$, $I(\Gamma_r) \subset I(\gamma)$. With our choice of γ_l and γ_r, an observer moving along L in the direction of increasing t finds the domain $G_l = I(\Gamma_l)$ on his left, and the domain $G_r = I(\Gamma_r)$ on his right, a fact anticipated by the notation. Therefore it is appropriate to write

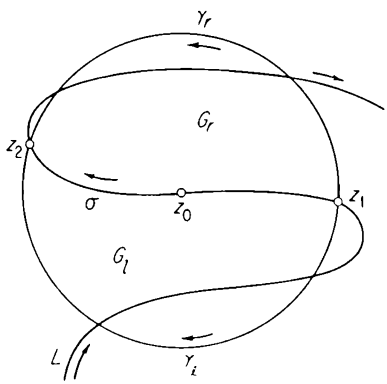

FIGURE 14.2

$$f_l(z_0) = \lim_{\substack{z \to z_0 \\ z \in G_l}} f(z), \qquad f_r(z_0) = \lim_{\substack{z \to z_0 \\ z \in G_r}} f(z).$$

We now derive explicit formulas for these two boundary values, and show that they satisfy formula (14.35).

Since $z \in G_l$ implies $z \in E(\Gamma_r)$, and since $\varphi(z)$ is analytic on $\overline{I(\Gamma_r)} \subset \overline{I(\gamma)}$, it follows from Theorem 14.1 that if $z \in G_l$, then

$$\frac{1}{2\pi i} \int_{\Gamma_r} \frac{\varphi(\zeta)}{\zeta - z}\, d\zeta = 0,$$

[11] Unless L is closed, we assume that z_0 is not one of the end points of L.

or

$$\frac{1}{2\pi i}\int_{\sigma}\frac{\varphi(\zeta)}{\zeta - z}\,d\zeta = \frac{1}{2\pi i}\int_{\gamma_r}\frac{\varphi(\zeta)}{\zeta - z}\,d\zeta. \tag{14.36}$$

Let L_1 be the part of L joining $z_a = \lambda(a)$ to z_1, and let L_2 be the part of L joining z_2 to $z_b = \lambda(b)$. Then $L = L_1 + \sigma + L_2$, and therefore, according to (14.36),

$$f(z) = \int_L \frac{\varphi(\zeta)}{\zeta - z}\,d\zeta = \int_{L_1+\sigma+L_2}\frac{\varphi(\zeta)}{\zeta - z}\,d\zeta = \int_{L_1+\gamma_r+L_2}\frac{\varphi(\zeta)}{\zeta - z}\,d\zeta \tag{14.37}$$

if $z \in G_l$. Since $L_1 + \gamma_r + L_2$ is a rectifiable curve (but perhaps not a Jordan curve), the right-hand side of (14.37) is an integral of the Cauchy type, which, according to Theorem 14.5, represents a function $\psi_l(z)$ such that $\psi_l(z)$ is analytic (and hence continuous) in any neighborhood $\mathcal{N}(z_0)$ containing no points of $L_1 + \gamma_r + L_2$. Moreover, as shown by (14.37), $\psi_l(z)$ coincides with $f(z)$ on $\mathcal{N}(z_0) \cap G_l$. It follows that

$$f_l(z_0) = \lim_{\substack{z \to z_0 \\ z \in G_l}} f(z) = \lim_{\substack{z \to z_0 \\ z \in G_l}} \psi_l(z) = \psi_l(z_0) = \frac{1}{2\pi i}\int_{L_1+\gamma_r+L_2}\frac{\varphi(\zeta)}{\zeta - z_0}\,d\zeta. \tag{14.38}$$

In exactly the same way, we find that

$$f(z) = \frac{1}{2\pi i}\int_{L_1+\gamma_l+L_2}\frac{\varphi(\zeta)}{\zeta - z}\,d\zeta$$

if $z \in G_r$, where the integral on the right represents a function $\psi_r(z)$ which is analytic in any neighborhood $\mathcal{N}(z_0)$ containing no points of $L_1 + \gamma_l + L_2$, and which coincides with $f(z)$ on $\mathcal{N}(z_0) \cap G_r$. It follows that

$$f_r(z_0) = \lim_{\substack{z \to z_0 \\ z \in G_r}} f(z) = \lim_{\substack{z \to z_0 \\ z \in G_r}} \psi_r(z) = \psi_r(z_0) = \frac{1}{2\pi i}\int_{L_1+\gamma_l+L_2}\frac{\varphi(\zeta)}{\zeta - z_0}\,d\zeta. \tag{14.39}$$

Subtracting (14.39) from (14.38), we find that

$$f_l(z_0) - f_r(z_0) = \frac{1}{2\pi i}\int_{\gamma_r-\gamma_l}\frac{\varphi(\zeta)}{\zeta - z_0}\,d\zeta = \frac{1}{2\pi i}\int_{\gamma}\frac{\varphi(\zeta)}{\zeta - z_0}\,d\zeta,$$

since $\gamma = \gamma_r - \gamma_l$ (see Figure 14.2). But since $\varphi(z)$ is analytic on $\overline{I(\gamma)}$, the integral on the right is a Cauchy integral, and therefore, according to Theorem 14.1,

$$f_l(z_0) - f_r(z_0) = \varphi(z_0),$$

i.e., (14.35) holds with a suitable definition of the boundary values of $f(z)$, and the proof is complete.

Example. If $\varphi(z) \equiv 1$ and if we choose L to be the segment $\alpha: -1 \leqslant x \leqslant 1$, $y = 0$, traversed in the direction of increasing x, then (14.32) reduces to the following integral of the Cauchy type:

$$f(z) = \frac{1}{2\pi i} \int_{-1}^{1} \frac{dx}{x - z}.$$

Let $z_0 \in L$ be the origin. Then, since $\varphi(z)$ is analytic everywhere, the only restriction on the choice of the circle γ with center at the point $z_0 = 0$ is that γ must contain both a point of α lying to the left of 0 and a point of α lying to the right of 0. Suppose we choose γ to be the unit circle $|z| = 1$, which obviously satisfies this requirement. Then the arc σ figuring in the proof of Theorem 14.11 coincides with the segment α, whereas the arcs γ_l and γ_r are the upper and lower unit semicircles, respectively (draw a figure). Applying (14.38) and (14.39) to this case, we find that

$$f_l(0) = \frac{1}{2\pi i} \int_{\gamma_r} \frac{d\zeta}{\zeta} = \frac{1}{2\pi i} \int_{\pi}^{2\pi} \frac{ie^{i\theta}\, d\theta}{e^{i\theta}} = \frac{1}{2}$$

and

$$f_r(0) = \frac{1}{2\pi i} \int_{\gamma_l} \frac{d\zeta}{\zeta} = \frac{1}{2\pi i} \int_{\pi}^{0} \frac{ie^{i\theta}\, d\theta}{e^{i\theta}} = -\frac{1}{2},$$

so that

$$f_l(0) - f_r(0) = \varphi(0) = 1,$$

in keeping with (14.35).

*74. The Plemelj Formulas

In Theorem 14.11 it is assumed that the function $\varphi(z)$ is analytic at every point of the path of integration L. However, in applications of the theory of the Cauchy type, one often encounters cases where this assumption is not satisfied. Therefore, in this section we shall derive another version of Theorem 14.11, this time relaxing the analyticity requirement on $\varphi(z)$; simultaneously, we shall deduce important formulas for the boundary values $f_l(z_0)$ and $f_r(z_0)$, which differ from (14.38) and (14.39). We begin by proving some preliminary results involving rectifiable Jordan curves.

LEMMA 1. *Any rectifiable Jordan curve L of length l can be represented by an equation of the form*

$$z = \lambda(s) \qquad (0 \leqslant s \leqslant l), \tag{14.40}$$

where s is the length of the arc of L joining the initial point $\lambda(0)$ to the variable point $\lambda(s)$.

Proof. Suppose L is originally specified by an equation of the form

$$z = h(t) \qquad (a \leqslant t \leqslant b),$$

where $h(t)$ is continuous on $[a, b]$, and let $s(t)$ be the function equal to the length of the arc of L joining the initial point $h(a)$ to the variable point $h(t)$ [$s(t)$ exists since L is rectifiable]. Then $s(t)$ is continuous and increasing (see Problem 12.19). Moreover, since L is a Jordan curve, $s(t)$ is strictly increasing and has a continuous strictly increasing inverse $t = t(s)$, defined on the interval $0 \leqslant s \leqslant l$ (see Problems 4.8 and 6.7). Therefore, according to the discussion on p. 60, L has an equivalent description in terms of the equation

$$z = h(t) = h[t(s)] \equiv \lambda(s) \qquad (0 \leqslant s \leqslant l),$$

with the arc length s as parameter.

LEMMA 2. *Let L be a piecewise smooth Jordan curve of length l, with equation* (14.40) *in terms of the arc length s. Then $\lambda'(s)$ exists and $|\lambda'(s)| = 1$ everywhere in $[0, l]$, except possibly at a finite number of points.*

Proof. If L is smooth, then $\lambda'(s)$ exists and is continuous for all $0 \leqslant s \leqslant l$ (why?), and moreover

$$s = \int_0^s |\lambda'(x)|\, dx \qquad (0 \leqslant s \leqslant l)$$

[cf. (12.6)]. Differentiating this formula with respect to s, we find that $|\lambda'(s)| = 1$ for all $0 \leqslant s \leqslant l$. The lemma now follows from the observation that a piecewise smooth curve can be decomposed into a finite number of smooth curves (see p. 247).

DEFINITION 1. *A set E of points on the real line is said to be of measure zero if given any $\varepsilon > 0$, there is a countable set of open intervals*

$$(a_1, b_1), \ldots, (a_n, b_n), \ldots \qquad (a_n < b_n),$$

depending on the choice of ε, such that

$$E \subset (a_1, b_1) \cup \cdots \cup (a_n, b_n) \cup \cdots$$

and

$$\sum_{n=1}^{\infty} (b_n - a_n) < \varepsilon.$$

LEMMA 2′. *Let L be a rectifiable Jordan curve of length l, with equation* (14.40) *in terms of the arc length s. Then $\lambda'(s)$ exists and $|\lambda'(s)| = 1$ for all s in $[0, l]$, except possibly on a set of points of measure zero.*

The proof of Lemma 2′ is beyond the scope of this book, but can be found, in one form or another, in many textbooks on real variable theory.[12]

---

[12] See e.g., C. Goffman, *Real Functions*, Holt, Rinehart and Winston, Inc., New York (1953), esp. Chap. 19, Secs. 7 and 8.

LEMMA 3. *Let L be a rectifiable Jordan curve of length l, with equation* (14.40) *in terms of the arc length s, and suppose $\lambda'(s_0)$ exists and is nonzero at some point $z_0 = \lambda(s_0) \in L$, so that L has a well-defined tangent at z_0. Let G_θ be the interior of an angle A of 2θ radians $(0 < \theta < \pi/2)$ with vertex at z_0 and bisector lying along the normal to L at z_0,*[13] *and let G_θ^* be the interior of the vertical angle of A (see Figure* 14.3). *Then there is a neighborhood $\mathcal{N}(z_0)$ such that*

$$(L \cap \mathcal{N}(z_0)) \cap (G_\theta \cup G_\theta^*) = 0,$$

i.e., any point of L lying sufficiently close to z_0 belongs to the complement of the set $G_\theta \cup G_\theta^*$.

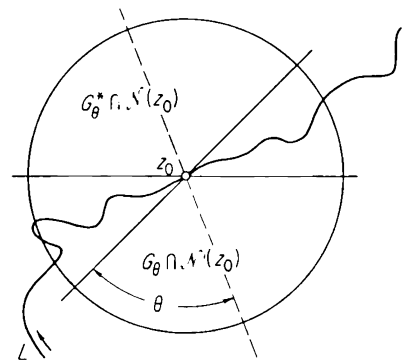

FIGURE 14.3

Proof. Suppose the lemma is false. Then there exists a sequence of points $z_n = \lambda(s_n) \in L$, $z_n \in G_\theta \cup G_\theta^*$, such that $z_n \to z_0 = \lambda(s_0)$ as $n \to \infty$. The mapping $z = \lambda(s)$, $0 \leqslant s \leqslant l$, being a one-to-one continuous mapping of a compact set,[14] is a homeomorphism (see Problem 6.5), and hence $s_n \to s_0$ as $n \to \infty$. Since

$$\lim_{n \to \infty} \frac{z_n - z_0}{s_n - s_0} = \lambda'(s_0) \neq 0,$$

the inclination of the vector $z_n - z_0$ must approach $\operatorname{Arg} \lambda'(s_0)$, for sufficiently large n, but this is impossible, since $z_n \in G_\theta \cup G_\theta^*$ for all n. This contradiction proves the lemma.

DEFINITION 2. *Let L, z_0, G_θ and G_θ^* be the same as in Lemma* 3. *Then a variable point z is said to approach z_0 along a nontangential path if $z \to z_0$ and $z \in G_\theta \cup G_\theta^*$, for some θ in the interval $(0, \pi/2)$.*[15]

We are now in a position to state and prove the new version of Theorem 14.11 alluded to earlier:

THEOREM 14.12. *Let L be a rectifiable Jordan curve (not necessarily closed) of length l, with equation*

$$z = \lambda(s) \qquad (0 \leqslant s \leqslant l)$$

---

[13] We define the *normal* to L at z_0 as the ray drawn from z_0 which is obtained when the tangent to L at z_0 is rotated 90° in the clockwise direction. This corresponds to the exterior normal in the case where L is closed.

[14] If L is a closed Jordan curve, consider instead the mapping $z = \lambda(s)$ on an interval $[0, l - \varepsilon]$ or $[\varepsilon, l]$ containing s_0 $(\varepsilon > 0)$.

[15] For example, $z \to z_0$ along a nontangential path if $z \to z_0$ along the normal to L at z_0.

in terms of the arc length s, and suppose $\lambda'(s_0)$ exists and is nonzero at some point $z_0 = \lambda(s_0) \in L$. Moreover, suppose $\varphi(z)$ satisfies an inequality of the form

$$|\varphi(z) - \varphi(z_0)| \leqslant C|z - z_0|^k \tag{14.41}$$

for all $z \in L$, where $C > 0$, $k > 0$, are constants. Then the integral of the Cauchy type

$$f(z) = \frac{1}{2\pi i}\int_L \frac{\varphi(\zeta)}{\zeta - z}\,d\zeta \tag{14.42}$$

satisfies the formula

$$f_l(z_0) - f_r(z_0) = \varphi(z_0), \tag{14.43}$$

with a suitable definition of the boundary values $f_l(z_0)$ and $f_r(z_0)$.[16]

Proof. First we prove that the improper integral

$$J_0 = \frac{1}{2\pi i}\int_L \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z_0}\,d\zeta \tag{14.44}$$

is absolutely convergent. Let $L_{\alpha\beta}$ be the part of the curve L joining the point $\lambda(\alpha)$ to the point $\lambda(\beta)$, where $s_0 < \alpha < \beta \leqslant b$, say. Then, according to Problem 12.20 and the inequality (14.41),

$$\left|\frac{1}{2\pi i}\int_{L_{\alpha\beta}} \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z_0}\,d\zeta\right| \leqslant \frac{C}{2\pi}\int_{L_{\alpha\beta}} |\zeta - z_0|^{k-1}\,ds$$

($|d\zeta| = ds$). By hypothesis,

$$\lim_{s \to s_0}\left|\frac{\zeta - z_0}{s - s_0}\right| = |\lambda'(s_0)| \neq 0,$$

where $\zeta \in L$, and hence there is a $\rho > 0$ such that

$$\left|\frac{\zeta - z_0}{s - s_0}\right| > \frac{1}{2}|\lambda'(s_0)|$$

if $|\zeta - z_0| < \rho$. Moreover,

$$\left|\frac{\zeta - z_0}{s - s_0}\right| > \frac{\rho}{l}$$

if $|\zeta - z_0| \geqslant \rho$. Therefore

$$\left|\frac{\zeta - z_0}{s - s_0}\right| > c > 0 \tag{14.45}$$

---

[16] Again, as in Theorem 14.11, we assume that z_0 is not one of the end points of L, where this assumption can be dropped if L is closed. Note that (14.41) implies that $\varphi(z)$ is continuous at z_0, but the converse is not true (see Problem 14.17).

for all $\zeta \in L$. On the other hand,

$$\left|\frac{\zeta - z_0}{s - s_0}\right| \leqslant 1 \tag{14.46}$$

for all $\zeta \in L$, by the very meaning of s. It follows that

$$\begin{aligned}\left|\frac{1}{2\pi i}\int_{L_{\alpha\beta}} \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z_0}\, d\zeta\right| &\leqslant C' \int_\alpha^\beta (s - s_0)^{k-1}\, ds \\ &= C'' \frac{(\beta - s_0)^k - (\alpha - s_0)^k}{k},\end{aligned} \tag{14.47}$$

where C' and C'' denote other positive constants.[17] Since the right-hand side of (14.47) becomes arbitrarily small when both α and β are sufficiently close to s_0, and since the same is true if we choose $a \leqslant \alpha < \beta < s_0$ instead of $s_0 < \alpha < \beta \leqslant b$, the integral J_0 is absolutely convergent, as asserted.[18]

Next we write the integral

$$f(z) = \frac{1}{2\pi i}\int_L \frac{\varphi(\zeta)}{\zeta - z}\, d\zeta$$

in the form

$$f(z) = \frac{1}{2\pi i}\int_L \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z}\, d\zeta + \frac{1}{2\pi i}\int_L \frac{\varphi(z_0)}{\zeta - z}\, d\zeta, \tag{14.48}$$

and show that

$$\lim_{z \to z_0} \frac{1}{2\pi i}\int_L \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z}\, d\zeta = J_0, \tag{14.49}$$

provided that z approaches z_0 *along a nontangential path* (see Definition 2). To this end, given any θ in the interval $(0, \pi/2)$, we first choose θ_0 such that $0 < \theta < \theta_0 < \pi/2$, and then we find a circle $\gamma: |z - z_0| = \rho$ such that

$$(L \cap I(\gamma)) \cap (G_{\theta_0} \cup G^*_{\theta_0}) = 0$$

(the existence of γ is guaranteed by Lemma 3), and such that γ contains at least one point $z = \lambda(s)$ with $0 \leqslant s < s_0$ and at least one point with $s_0 < s \leqslant l$. Let the points z_1, z_2 and the arcs σ, L_1, L_2, γ_l, γ_r have

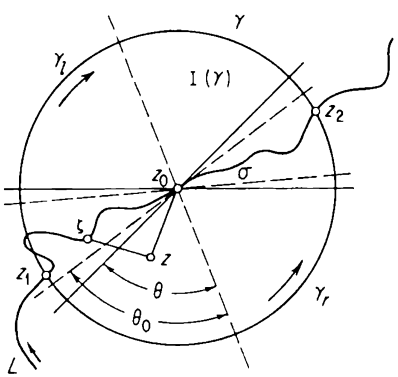

FIGURE 14.4

[17] Note that we use (14.45) if $k < 1$ and (14.46) if $k > 1$ (see Problem 14.18).

[18] Here we use the analogue of the Cauchy convergence criterion (Theorem 3.4) for improper integrals.

exactly the same meaning as in the proof of Theorem 14.11, where the parameter t is now the arc length. However, it must now be kept in mind that all these quantities, like γ itself, depend on the radius ρ, which will be varied in the course of the proof. As shown by Figure 14.4,

$$\left|\frac{z_0 - z}{\zeta - z}\right| < \csc(0_0 - 0)$$

for any points $\zeta \in \sigma$, $z \in I(\gamma) \cap (G_0 \cup G_0^*)$, and therefore

$$K = \left|J_0 - \frac{1}{2\pi i}\int_L \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z}\,d\zeta\right|$$
$$= \left|\frac{1}{2\pi i}\int_L [\varphi(\zeta) - \varphi(z_0)]\frac{z_0 - z}{(\zeta - z)(\zeta - z_0)}\,d\zeta\right| \leqslant J_1 + J_2,$$

where

$$J_1 = \left|\frac{1}{2\pi i}\int_{L'} [\varphi(\zeta) - \varphi(z_0)]\frac{z_0 - z}{(\zeta - z)(\zeta - z_0)}\,d\zeta\right|,$$
$$J_2 = \csc(0_0 - 0)\frac{1}{2\pi}\left|\int_\sigma \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z_0}\right| ds,$$

and $L' = L_1 \cup L_2$, i.e., L' is the curve L with the arc σ deleted. Given any $\varepsilon > 0$, we can choose ρ so small that $J_2 < \varepsilon/2$ (note that the integral in J_2 converges and the length of σ decreases with ρ). Having chosen such a value of ρ, we observe that

$$|\zeta - z_0| \geqslant \delta > 0$$

if $\zeta \in L'$ and δ is the distance between z_0 and L', whereas

$$|\zeta - z| \geqslant \delta - |z_0 - z| > \frac{\delta}{2}$$

if $\zeta \in L'$ and $|z - z_0| < \delta/2$. These two inequalities imply that

$$J_1 < \frac{|z_0 - z|}{\pi\delta^2}\int_{L'} |\varphi(\zeta) - \varphi(z_0)|\,ds \leqslant \frac{|z_0 - z|}{\pi\delta^2}\int_L |\varphi(\zeta) - \varphi(z_0)|\,ds,$$

so that $J_1 < \varepsilon/2$ for all z sufficiently close to z_0. It follows that

$$K \leqslant J_1 + J_2 < \varepsilon$$

for all $z \in G_0 \cup G_0^*$ sufficiently close to z_0. Since ε is arbitrary, this proves (14.49), provided that z approaches z_0 along a nontangential path.

We now consider the second integral in the right-hand side of (14.48). Obviously, this is an integral of the Cauchy type (14.42), where the function in the numerator of the integrand is the constant $\varphi(z_0)$ and is therefore analytic. Moreover, it is easy to see that[19]

$$\psi(z) \equiv \frac{1}{2\pi i}\int_L \frac{\varphi(z_0)}{\zeta - z}\,d\zeta = \frac{\varphi(z_0)}{2\pi i}\operatorname*{Var}_{\zeta \in L} \operatorname{Ln}(\zeta - z),$$

[19] By $\operatorname*{Var}_{\zeta \in L} h(\zeta)$ is meant the change in $h(\zeta)$ as ζ traverses the curve L.

but this fact will not be used here. Instead, we have recourse to the following formulas proved in Sec. 73 [cf. (14.38) and (14.39)]:

$$\psi_l(z_0) = \frac{1}{2\pi i}\int_{L_1+\gamma_r+L_2} \frac{\varphi(z_0)}{\zeta - z_0}\,d\zeta, \qquad \psi_r(z_0) = \frac{1}{2\pi i}\int_{L_1+\gamma_l+L_2} \frac{\varphi(z_0)}{\zeta - z_0}\,d\zeta. \tag{14.50}$$

A comparison of (14.48), (14.49) and (14.50) shows that

$$f_l(z_0) = \frac{1}{2\pi i}\int_L \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z_0}\,d\zeta + \frac{1}{2\pi i}\int_{L_1+\gamma_r+L_2} \frac{\varphi(z_0)}{\zeta - z_0}\,d\zeta, \tag{14.51}$$

$$f_r(z_0) = \frac{1}{2\pi i}\int_L \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z_0}\,d\zeta + \frac{1}{2\pi i}\int_{L_1+\gamma_l+L_2} \frac{\varphi(z_0)}{\zeta - z_0}\,d\zeta. \tag{14.52}$$

Subtracting (14.52) from (14.51), we find that

$$f_l(z_0) - f_r(z_0) = \frac{1}{2\pi i}\int_{|\zeta - z_0| = \rho} \frac{\varphi(z_0)}{\zeta - z_0}\,d\zeta = \varphi(z_0),$$

which agrees with formula (14.43). It should be noted that we have imposed no analyticity requirement on the function $\varphi(z)$.[20]

Remark 1. We can derive simpler expressions for the boundary values $f_l(z_0)$ and $f_r(z_0)$. For example, (14.51) can be written in the form

$$\begin{aligned} f_l(z_0) &= \frac{1}{2\pi i}\int_{L'} \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z_0}\,d\zeta + \frac{1}{2\pi i}\int_{\sigma} \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z_0} \\ &\quad + \frac{1}{2\pi i}\int_{L'} \frac{\varphi(z_0)}{\zeta - z_0}\,d\zeta + \frac{1}{2\pi i}\int_{\gamma_r} \frac{\varphi(z_0)}{\zeta - z_0}\,d\zeta \\ &= \frac{1}{2\pi i}\int_{L'} \frac{\varphi(\zeta)}{\zeta - z_0}\,d\zeta + \frac{1}{2\pi i}\int_{\sigma} \frac{\varphi(\zeta) - \varphi(z_0)}{\zeta - z_0}\,d\zeta \\ &\quad + \frac{1}{2\pi i}\int_{\gamma_r} \frac{\varphi(z_0)}{\zeta - z_0}\,d\zeta. \end{aligned} \tag{14.53}$$

We again remind the reader that L', σ and γ_r implicitly involve the radius ρ of the circle $\gamma: |z - z_0| = \rho$. As $\rho \to 0$, the second integral on the right approaches zero, since the integral (14.44) converges. The third integral on the right can be written as

$$\frac{\varphi(z_0)}{2\pi i} \operatorname*{Var}_{\zeta \in \gamma_r} \operatorname{Ln}(\zeta - z_0) = \frac{\varphi(z_0)}{2\pi} \operatorname*{Var}_{\zeta \in \gamma_r} \operatorname{Arg}(\zeta - z_0) = \frac{\varphi(z_0)}{2\pi\rho} \text{ Length of } \gamma_r,$$

since $\operatorname{Ln}(z - z_0) = \ln \rho + i \operatorname{Arg}(z - z_0)$ if $z \in \gamma$. But as already shown (cf. Lemma 3), the arc $\sigma \subset L$ can be included in a pair of arbitrarily narrow vertical angles with common vertex at z_0 and bisector lying along the tangent

[20] As in the proof of Theorem 14.11, the limit $f_l(z_0)$ is evaluated for $z \in I(\sigma - \gamma_l)$, but now z must approach z_0 along a nontangential path, and similarly for $f_r(z_0)$.

to L at z_0. Therefore, as $\rho \to 0$, the arc γ_r approaches one of the semicircles into which this tangent divides the circle γ, and therefore

$$\lim_{\rho \to 0} \int_{\gamma_r} \frac{\varphi(z_0)}{\zeta - z_0}\, d\zeta = \frac{\varphi(z_0)}{2\pi}\, \pi = \frac{1}{2}\, \varphi(z_0).$$

Since the left-hand side of (14.53) is independent of ρ, we conclude that the limit

$$\lim_{\rho \to 0} \frac{1}{2\pi i} \int_{L'} \frac{\varphi(\zeta)}{\zeta - z_0}\, d\zeta, \tag{14.54}$$

which we denote by

$$\frac{1}{2\pi i} \int_L' \frac{\varphi(\zeta)}{\zeta - z_0}\, d\zeta,$$

also exists. Combining all these facts, we finally find that

$$f_i(z_0) = \frac{1}{2\pi i} \int_L' \frac{\varphi(\zeta)}{\zeta - z_0}\, d\zeta + \frac{1}{2}\, \varphi(z_0), \tag{14.55}$$

which together with (14.43) implies

$$f_r(z_0) = \frac{1}{2\pi i} \int_L' \frac{\varphi(\zeta)}{\zeta - z_0}\, d\zeta - \frac{1}{2}\, \varphi(z_0). \tag{14.56}$$

Formulas (14.55) and (14.56) have important applications to mechanics, and are called the *Plemelj formulas* (or, less often, the *Sokhotski formulas*).[21]

Remark 2. It should be noted that the integral

$$\frac{1}{2\pi i} \int_L \frac{\varphi(\zeta)}{\zeta - z_0}\, d\zeta \tag{14.57}$$

may well be divergent. We have merely proved the existence of the limit (14.54), known as the *Cauchy principal value* of (14.57).

PROBLEMS

In Problems 14.1 through 14.6, it is assumed that L is a closed rectifiable Jordan curve, with interior $I(L)$ and exterior $E(L)$, traversed in the counterclockwise direction.

---

[21] The basic results of this and the preceding sections were first obtained in 1873 by the Russian mathematician Y. V. Sokhotski. Refinements were subsequently introduced by J. Plemelj, I. I. Privalov and V. V. Golubyev. In particular, see I. I. Privalov, *Randeigenschaften Analytischer Funktionen*, second revised edition (edited by A. I. Markushevich), VEB Deutscher Verlag der Wissenschaften, Berlin (1956), and F. D. Gakhov, *Boundary Value Problems*, second edition (in Russian), Gos. Izd. Fiz.-Mat. Lit., Moscow (1963). For applications of integrals of the Cauchy type to problems of mechanics, especially in the theory of elasticity, see N. I. Muskhelishvili, *Singular Integral Equations*, second edition (translated by J. R. M. Radok), P. Noordhoff, Groningen (1953).

14.1. Evaluate the integral

$$\int_L \frac{dz}{z^2 + 9},$$

given that

a) $I(L)$ contains the point $3i$, while $E(L)$ contains the point $-3i$;
b) $I(L)$ contains the point $-3i$, while $E(L)$ contains the point $3i$;
c) $I(L)$ contains both points $\pm 3i$.

14.2. Assuming that none of the points 0, 1 and -1 lies on L, calculate all possible values of the integral

$$\int_L \frac{dz}{z(z^2 - 1)}$$

for various positions of L.

14.3. Evaluate the integral

$$\int_{|z-a|=a} \frac{z}{z^4 - 1}\, dz \qquad (a > 1).$$

14.4. Evaluate the integral

$$\frac{1}{2\pi i} \int_L \frac{e^z}{z^2 + a^2}\, dz,$$

if $I(L)$ contains the closed disk $|z| \leqslant a$.

14.5. Evaluate the integral

$$\frac{1}{2\pi i} \int_L \frac{ze^z}{(z - a)^3}\, dz,$$

given that $a \in I(L)$.

Hint. Use formula (14.20).

14.6. Let $f(z)$ be a single-valued analytic function on a domain G which contains a contour L and its interior $I(L)$. Prove that if $I(L)$ contains the point $z = 0$, then

$$\frac{1}{2\pi i} \int_L f'(z) \operatorname{Ln} z\, dz = f(z_0) - f(0),$$

where Ln z denotes any single-valued branch of the logarithm, and z_0 is the initial (and final) point of integration.

Hint. Use Problem 13.19.

14.7. Evaluate the integral

$$\frac{1}{2\pi i} \int_L z^2 \operatorname{Ln} \frac{z + 1}{z - 1}\, dz,$$

where Ln $a = \ln a$ for $a > 0$, given that

a) L is the circle $|z| = 2$;
b) L is the circle $|z - 1| = 1$ and the initial point of integration is $z_0 = 1 + i$.

Hint. Use the preceding problem.

Ans. b) $1 - \frac{2i}{3}$.

14.8. Prove that

$$\int_0^{2\pi} \cos(\cos\theta)\cosh(\sin\theta)\,d\theta = 2\pi.$$

Hint. Apply formula (14.5) to the function $f(z) = \cos z$.

14.9. Prove that if $f(z)$ is a nonvanishing analytic function on a domain G, and if G contains the circle $\gamma_\rho: |z - z_0| = \rho$ and its interior $I(\gamma_\rho)$, then

$$\ln|f(z_0)| = \frac{1}{2\pi}\int_0^{2\pi} \ln|f(z_0 + \rho e^{i\theta})|\,d\theta.$$

14.10. Prove that

$$\int_0^{2\pi} \ln[\cosh^2(\sin\theta) - \sin^2(\cos\theta)]\,d\theta = 0.$$

Hint. Apply the preceding problem to the function $f(z) = \cos z$.

14.11. Prove that if the function $f(z)$ is one-to-one and continuous on a domain G, then $|f(z)|$ cannot have a relative maximum in the wide sense at any point of G.

Hint. Use Theorem 6.1.

14.12. Let $f(z)$ be an analytic function on a domain G, and, as usual, let

$$M(\rho) = \max_{|z - z_0| = \rho} |f(z)|,$$

where $z_0 \in G$. Prove that $M(\rho)$ is a continuous function in the interval $(0, \Delta)$, where Δ is the distance between z_0 and the boundary of G.

Comment. We have already proved in Theorem 14.4, Corollary 2 that $M(\rho)$ is strictly increasing in the interval $(0, \Delta)$.

14.13. Is Theorem 14.5 applicable to either of the integrals

$$\int_{|\zeta|=1} \frac{\bar{\zeta}}{\zeta - z}\,d\zeta, \quad \int_{|\zeta|=1} \frac{\zeta^{-1}}{\zeta - z}\,d\zeta \qquad (|z| \neq 1)?$$

14.14. Let L be a closed rectifiable Jordan curve, traversed in the counterclockwise direction, let $f(z)$ be analytic on $E(L)$,[22] and suppose that

$$\lim_{z \to \infty} f(z) = A.$$

Prove that

$$\frac{1}{2\pi i}\int_L \frac{f(\zeta)}{\zeta - z}\,d\zeta = A \qquad \text{if} \quad z \in I(L),$$

and

$$\frac{1}{2\pi i}\int_L \frac{f(\zeta)}{\zeta - z}\,d\zeta = -f(z) + A \qquad \text{if} \quad z \in E(L).$$

Hint. Make the transformation $w = 1/z$, $\eta = 1/\zeta$.

---

[22] In particular, $f(z)$ is analytic at infinity (see Remark 1, p. 125).

Comment. This result is sometimes called *Cauchy's integral formula for an unbounded domain.*

14.15. A real function $f(x, y)$ of two real variables x and y is said to be *harmonic* on a domain G if it has continuous first and second partial derivatives on G, and if

$$\frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} = 0$$

at every point of G. Prove that if $f(z)$ is analytic on G, then its real and imaginary parts are harmonic on G.

14.16. As in footnote 9, p. 303, let $\{a_n\}$ be a bounded sequence of real numbers, and let $\overline{\lim_{n \to \infty}} a_n = A$ be the largest limit point of $\{a_n\}$. Prove that

a) Given any $\varepsilon > 0$, there is an integer $N(\varepsilon) > 0$ such that

$$a_n < A + \varepsilon$$

for all $n > N(\varepsilon)$;

b) Given any $\varepsilon > 0$ and any integer $m > 0$, there is an integer $n > m$ such that

$$a_n > A - \varepsilon.$$

State and prove the analogue of this result for the *smallest* limit point of $\{a_n\}$, called the *limit inferior* or *lower limit* of $\{a_n\}$, and denoted by $\underline{\lim_{n \to \infty}} a_n$.

14.17. Give an example of a continuous function $\varphi(z)$ defined on a curve L which satisfies no inequality of the form

$$|\varphi(z) - \varphi(z_0)| \leqslant C\,|z - z_0|^k \qquad (z \in L),$$

regardless of the choice of $C > 0$, $k > 0$.

Hint. Consider the function

$$\varphi(x) = \begin{cases} \dfrac{1}{\ln x} & \text{if} \quad 0 < x \leqslant \tfrac{1}{2}, \\ 0 & \text{if} \quad -\tfrac{1}{2} \leqslant x \leqslant 0. \end{cases}$$

14.18. Let L be a continuous curve, and suppose that

$$|\varphi(z) - \varphi(z')| \leqslant C\,|z - z'|^k$$

for all z, z' in L, where $C > 0$, $k > 0$. Prove that $\varphi(z) \equiv$ const on L if $k > 1$.

In Problems 14.19 through 14.22, $f(z)$ denotes an integral of the Cauchy type, with density $\varphi(\zeta)$, i.e.,

$$f(z) = \frac{1}{2\pi i} \int_L \frac{\varphi(\zeta)}{\zeta - z}\, d\zeta,$$

where L is a rectifiable Jordan curve. For simplicity, we shall consistently use ζ to denote a variable point of L, reserving z for a variable point which does not belong to L. Thus, the boundary values defined in Secs. 73 and 74 will be denoted by $f_i(\zeta)$ and $f_r(\zeta)$.

14.19. Verify the formula

$$f_i(\zeta) - f_r(\zeta) = \varphi(\zeta)$$

for the case where L is the unit circle $|\zeta| = 1$ and

$$\varphi(\zeta) = \frac{2}{\zeta(\zeta - 2)}.$$

Hint. Use Problem 14.14.

14.20. Find $f_i(\zeta)$ and $f_r(\zeta)$ for each of the following densities, where L is closed and n is a positive integer:

a) $\varphi(\zeta) = (\zeta - a)^n$; b) $\varphi(\zeta) = \dfrac{1}{(\zeta - a)^n}$, $a \in I(L)$;

c) $\varphi(\zeta) = \dfrac{1}{(\zeta - a)^n}$, $a \in E(L)$.

Ans. b) $f_i(\zeta) \equiv 0, f_r(\zeta) = -\dfrac{1}{(\zeta - a)^n}$.

14.21. Find $f_i(\zeta)$ and $f_r(\zeta)$ when $\varphi(\zeta) \equiv 1$, and L is the segment $-1 \leqslant x \leqslant 1$, $y = 0$, traversed in the direction of increasing x.

Ans. $f_i(\zeta) = \dfrac{1}{2\pi i} \ln \dfrac{1 - \zeta}{1 + \zeta} + \dfrac{1}{2}, \qquad f_r(\zeta) = \dfrac{1}{2\pi i} \ln \dfrac{1 - \zeta}{1 + \zeta} - \dfrac{1}{2}.$

Comment. The case $\zeta = 0$ has already been discussed in the example on p. 309.

14.22. Let L be the semicircle $|\zeta| = R$, $0 \leqslant \arg \zeta \leqslant \pi$, with initial point R, and let $\varphi(\zeta) \equiv 1$. Find $f(z)$, $f_i(\zeta)$ and $f_r(\zeta)$. In particular, calculate $f(0)$, $f_i(iR)$ and $f_r(iR)$.

Ans. $f(0) = \frac{1}{2}, f_i(iR) = \frac{3}{4}, f_r(iR) = -\frac{1}{4}$.

14.23. If $n > 0$ is an odd integer, prove the existence of the Cauchy principal value of the improper integral

$$\int_a^b \frac{dx}{(x - c)^n},$$

where $a < c < b$.

14.24. Find the Cauchy principal value of the improper integral

$$\frac{1}{\pi i} \int_{|z| = 1} \frac{z^2}{z^3 - 1}\, dz.$$

Ans. 1.

15

UNIFORM CONVERGENCE. INFINITE PRODUCTS

75. Uniformly Convergent Series

Let

$$\sum_{n=1}^{\infty} f_n(z) = f_1(z) + \cdots + f_n(z) + \cdots \tag{15.1}$$

be an (infinite) series whose terms are functions of a complex variable defined on a set E in the complex plane, and let

$$s_n(z) = f_1(z) + \cdots + f_n(z) \tag{15.2}$$

be the nth partial sum of (15.1). Then the series (15.1) is said to be *uniformly convergent on* E, if given any $\varepsilon > 0$, there exists an integer $N(\varepsilon) > 0$ such that

$$|s_{n+p}(z) - s_n(z)| < \varepsilon$$

for all $n > N(\varepsilon)$, all $p > 0$ and all $z \in E$.[1] It follows from the Cauchy convergence criterion (Theorem 3.4) that every series which is uniformly convergent on a set E is also convergent on E, i.e., the function

$$f(z) = \lim_{n \to \infty} s_n(z),$$

[1] For brevity, we shall usually just say "for all $n > N(\varepsilon)$, $p > 0$ and $z \in E$." It may sometimes be appropriate to write $N(E, \varepsilon)$ instead of $N(\varepsilon)$, to emphasize the fact that $N(E, \varepsilon)$ depends on the set E (which may not be fixed), as well as on the number ε (cf. the remark on p. 325).

called the *sum* of the series (15.1) exists for all $z \in E$. However, a series can be convergent on E without being uniformly convergent on E (see example below).

Remark. In some problems, especially those involving power series, we find it convenient to introduce an extra "zeroth" term $f_0(z)$, writing

$$\sum_{n=0}^{\infty} f_n(z) = f_0(z) + f_1(z) + \cdots + f_n(z) + \cdots$$

and

$$s_n(z) = f_0(z) + f_1(z) + \cdots + f_n(z),$$

instead of (15.1) and (15.2).

Example. The *geometric series*

$$\sum_{n=0}^{\infty} z^n = 1 + z + \cdots + z^n + \cdots \tag{15.3}$$

converges on the open unit disk $K: |z| < 1$, since

$$1 + z + \cdots + z^n = \frac{1 - z^{n+1}}{1 - z} \to \frac{1}{1 - z} \quad \text{as} \quad n \to \infty$$

if $|z| < 1$. Moreover,

$$\begin{aligned} |s_{n+p}(z) - s_n(z)| &= |z^{n+1}(1 + z + \cdots + z^{p-1})| \\ &= \frac{|z|^{n+1}|1 - z^p|}{|1 - z|} \geqslant \frac{|z|^{n+1}(1 - |z|^p)}{|1 - z|}, \end{aligned}$$

and hence choosing

$$p = n, \quad z_n = \frac{n - 1}{n} \qquad (n > 0),$$

we find that[2]

$$|s_{2n}(z_n) - s_n(z_n)| \geqslant n\left(1 - \frac{1}{n}\right)^{n+1}\left[1 - \left(1 - \frac{1}{n}\right)^n\right] \to \infty \qquad \text{as} \quad n \to \infty.$$

Thus, for sufficiently large n and appropriate p and $z \in K$, we can make the difference $|s_{n+p}(z) - s_n(z)|$ as large as we please. It follows that the series (15.3) is not uniformly convergent on K.

THEOREM 15.1. *A convergent series*

$$\sum_{n=1}^{\infty} f_n(z) = f(z) \qquad (z \in E) \tag{15.4}$$

---

[2] Recall that $\lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^n = \frac{1}{e}$.

is uniformly convergent on E if and only if given any $\varepsilon > 0$, there exists an integer $N(\varepsilon) > 0$ such that

$$|s_n(z) - f(z)| < \varepsilon \tag{15.5}$$

for all $n > N(\varepsilon)$ and $z \in E$.

Proof. If the series (15.4) is uniformly convergent on E, then, given any $\varepsilon > 0$, there is an integer $N_1(\varepsilon) > 0$ such that

$$|s_{n+p}(z) - s_n(z)| < \varepsilon \tag{15.6}$$

for all $n > N_1(\varepsilon)$, $p > 0$ and $z \in E$. Taking the limit of (15.6) as $p \to \infty$, we find that

$$|f(z) - s_n(z)| < \varepsilon$$

for all $n > N_1(\varepsilon)$ and $z \in E$. Conversely, if there is an integer $N_2(\varepsilon) > 0$ such that

$$|s_n(z) - f(z)| < \frac{\varepsilon}{2}$$

for all $n > N_2(\varepsilon)$ and $z \in E$, then

$$\begin{aligned} |s_{n+p}(z) - s_n(z)| &= |[s_{n+p}(z) - f(z)] + [f(z) - s_n(z)]| \\ &\leqslant |s_{n+p}(z) - f(z)| + |s_n(z) - f(z)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

for all $n > N_2(\varepsilon)$, $p > 0$ and $z \in E$.

Remark. Thus, we could just as well have defined a uniformly convergent series on E as a series which converges on E to a sum $f(z)$ satisfying the condition (15.5).

Next we give an important sufficient condition for uniform convergence:

THEOREM 15.2 (*Weierstrass' M-test*). *Given a convergent series*

$$\sum_{n=1}^{\infty} M_n, \tag{15.7}$$

whose terms are nonnegative constants, suppose the functions $f_1(z), \ldots, f_n(z), \ldots$ are such that

$$|f_n(z)| \leqslant M_n \tag{15.8}$$

for all $z \in E$ and all n exceeding a certain integer $N > 0$. Then the series

$$\sum_{n=1}^{\infty} f_n(z) = f(z) \qquad (z \in E) \tag{15.9}$$

is uniformly convergent on E.

Proof. Since (15.7) converges, given any $\varepsilon > 0$, there is an integer $N_1(\varepsilon) > 0$ such that

$$M_{n+1} + \cdots + M_{n+p} < \varepsilon$$

for all $n > N_1(\varepsilon)$ and $p > 0$. If $N_2(\varepsilon) = \max\{N_1(\varepsilon), N\}$, then, according to (15.8),

$$\begin{aligned} |s_{n+p}(z) - s_n(z)| &\leqslant |f_{n+1}(z)| + \cdots + |f_{n+p}(z)| \\ &\leqslant M_{n+1} + \cdots + M_{n+p} < \varepsilon \end{aligned}$$

for all $n > N_2(\varepsilon)$, $p > 0$ and $z \in E$, i.e., (15.9) is uniformly convergent on E. We note, in passing, that (15.9) is also absolutely convergent on E.

Suppose every point of a set E is a limit point of E. Then E is said to be *dense in itself.* For example, any nonempty open set (in particular, a domain), or any continuous curve, is dense in itself.

THEOREM 15.3. *Given a uniformly convergent series* (15.9) *defined on a set E which is dense in itself, suppose every term $f_n(z)$, $n = 1, 2, \ldots$, is continuous on E. Then the sum $f(z)$ is also continuous on E.*

Proof. If z and z_0 are any two points of E, then

$$\begin{aligned} |f(z) - f(z_0)| &= |[f(z) - s_n(z)] + [s_n(z) - s_n(z_0)] + [s_n(z_0) - f(z_0)]| \\ &\leqslant |s_n(z) - f(z)| + |s_n(z) - s_n(z_0)| + |s_n(z_0) - f(z_0)|. \end{aligned}$$

Since (15.9) is uniformly convergent on E, given any $\varepsilon > 0$, there is an integer $N(\varepsilon) > 0$ such that

$$|s_n(z) - f(z)| < \frac{\varepsilon}{3},$$

for all $n > N(\varepsilon)$ and $z \in E$ (see Theorem 15.1), and in particular,

$$|s_n(z_0) - f(z_0)| < \frac{\varepsilon}{3}$$

for all $n > N(\varepsilon)$. Therefore, if n_0 is any integer greater than $N(\varepsilon)$, we can set $n = n_0$ and then combine the last three inequalities, obtaining

$$|f(z) - f(z_0)| < |s_{n_0}(z) - s_{n_0}(z_0)| + \frac{2\varepsilon}{3}. \tag{15.10}$$

Moreover, since $s_{n_0}(z)$ is continuous at $z_0 \in E$, being the sum of a *finite* number of functions continuous at z_0, we can choose $\delta(\varepsilon) > 0$ such that

$$|s_{n_0}(z) - s_{n_0}(z_0)| < \frac{\varepsilon}{3}$$

if $|z - z_0| < \delta$, and hence, according to (15.10),

$$|f(z) - f(z_0)| < \varepsilon$$

if $|z - z_0| < \delta$. But this is precisely what is meant by saying that $f(z)$ is continuous at z_0.

THEOREM 15.4. *Given a rectifiable curve L, suppose the series*

$$\sum_{n=1}^{\infty} f_n(z) = f(z) \tag{15.11}$$

is uniformly convergent on L, and suppose every term $f_n(z)$ is continuous on L. Then (15.11) *can be integrated term by term along L, i.e.,*

$$\sum_{n=1}^{\infty} \int_L f_n(z)\,dz = \int_L f(z)\,dz. \tag{15.12}$$

Proof. According to Theorem 15.3, $f(z)$ is continuous and hence integrable along L. Since (15.11) is uniformly convergent on L, given any $\varepsilon > 0$, there is an integer $N(\varepsilon) > 0$ such that

$$|s_n(z) - f(z)| < \frac{\varepsilon}{l}$$

for all $n > N(\varepsilon)$ and $z \in L$, where l is the length of L (see Theorem 15.1). Then obviously

$$\left| \sum_{k=1}^{n} \int_L f_k(z)\,dz - \int_L f(z)\,dz \right| = \left| \int_L [s_n(z) - f(z)]\,dz \right| < \frac{\varepsilon}{l}\, l = \varepsilon,$$

and since ε is arbitrary, the theorem is proved.

COROLLARY. *Given a rectifiable curve L contained in a domain G, suppose the series* (15.11) *is uniformly convergent on every compact subset of G, and suppose every term $f_n(z)$ is continuous on G. Then* (15.11) *can be integrated term by term along L.*

Proof. We need only note that any continuous curve (rectifiable or not) contained in G is a compact subset of G.

Remark. Obviously, every series which is uniformly convergent on a domain G is uniformly convergent on every compact subset of G.[3] However, the converse is not true. For example, consider the geometric series

$$\sum_{n=0}^{\infty} z^n = 1 + z + \cdots + z^n + \cdots,$$

which, as we have already shown (see p. 322) is convergent on the disk $K: |z| < 1$. Let F be any compact subset of K, and let $\delta > 0$ be the distance between F and the circle $|z| = 1$ (the boundary of K). Then for any point

[3] The locution "on compact subsets (of G)" is often encountered, instead of "on every compact subset of G." Later, we shall introduce the even terser phrase "inside G" (see p. 411).

$z \in F$, we have $|z| \leqslant 1 - \delta$, and hence

$$|z^{n+1} + \cdots + z^{n+p}| = \left| z^{n+1} \frac{1 - z^p}{1 - z} \right| \leqslant |z|^{n+1} \frac{1 + |z|^p}{1 - |z|} \leqslant (1 - \delta)^{n+1} \frac{2}{\delta}. \tag{15.13}$$

Since $(1 - \delta)^{n+1}(2/\delta) \to 0$ as $n \to \infty$, then, given any $\varepsilon > 0$, there is an integer $N(F, \varepsilon) > 0$ such that the left-hand side of (15.13) is less than ε for all $n > N(F, \varepsilon)$ and $z \in F$. It follows that the geometric series is uniformly convergent on every compact subset of K (but not on K itself, as already noted).

THEOREM 15.5. *A series*

$$\sum_{n=1}^{\infty} f_n(z) = f(z) \tag{15.14}$$

which is convergent on a domain G is uniformly convergent on every compact subset of G if and only if, given any point $z_0 \in G$, there is a neighborhood $\mathcal{N}(z_0) \subset G$ in which (15.14) *is uniformly convergent.*

Proof. The condition is obviously necessary, since given any $z_0 \in G$, (15.14) is uniformly convergent in some closed neighborhood

$$\overline{\mathcal{N}(z_0)}: |z - z_0| \leqslant \rho$$

contained in G, and hence uniformly convergent on $\mathcal{N}(z_0)$. To prove that the condition is sufficient, suppose there is a compact set $F \subset G$ on which (15.14) is not uniformly convergent, although the condition of the theorem is satisfied. Then there must exist a number $\varepsilon > 0$, a sequence $\{z_n\}$ of points in F, and a sequence $\{k_n\}$ of positive integers, where $k_1 < \cdots < k_n < \cdots$, such that

$$|s_{k_n}(z_n) - f(z_n)| \geqslant \varepsilon \tag{15.15}$$

for all $n = 1, 2, \ldots$. [This is the negative of the assertion that (15.14) is uniformly convergent on F.] The sequence $\{z_n\}$, being bounded, has at least one limit point ζ (see Theorem 3.1), and moreover contains a subsequence $\{z_n'\}$ converging to ζ. Since F is closed, $\zeta \in F \subset G$, and hence, by hypothesis, there is a neighborhood $\mathcal{N}(\zeta) \subset G$ in which the series (15.14) is uniformly convergent, i.e.,

$$|s_n(z) - f(z)| < \varepsilon$$

for sufficiently large n and all $z \in \mathcal{N}(\zeta)$. On the other hand, for sufficiently large n, $\mathcal{N}(\zeta)$ must contain points (in fact, all the points of $\{z_n'\}$ starting from a certain value of n) satisfying the opposite inequality (15.15). This contradiction completes the proof.

THEOREM 15.6 (*Weierstrass' theorem on uniformly convergent series of analytic functions*). *If the series*

$$\sum_{n=1}^{\infty} f_n(z) = f(z) \tag{15.16}$$

is uniformly convergent on every compact subset of a domain G, and if every term $f_n(z)$ is analytic on G, then the sum $f(z)$ of the series is also analytic on G. Moreover, the series (15.16) *can be differentiated term by term any number of times, i.e.,*[4]

$$\sum_{n=1}^{\infty} f_n^{(k)}(z) = f^{(k)}(z) \qquad (k = 1, 2, \ldots) \tag{15.17}$$

for all $z \in G$, and each differentiated series is uniformly convergent on every compact subset of G.

Proof. Let z_0 be an arbitrary point of G, and choose $\rho > 0$ such that G contains the circle $\gamma_\rho: |z - z_0| = \rho$ and its interior $I(\gamma_\rho)$. By hypothesis, the series (15.16) is uniformly convergent on γ_ρ, and hence so is each of the series

$$\frac{k!}{2\pi i} \sum_{n=1}^{\infty} \frac{f_n(z)}{(z - z_0)^{k+1}} = \frac{k!}{2\pi i} \frac{f(z)}{(z - z_0)^{k+1}} \qquad (k = 0, 1, 2, \ldots) \tag{15.18}$$

$(0! = 1)$, since

$$\left| \frac{k!}{2\pi i} \frac{1}{(z - z_0)^{k+1}} \right| = \frac{k!}{2\pi \rho^{k+1}}$$

for all $z \in \gamma_\rho$ (see Problem 15.3). Therefore we can integrate (15.18) term by term along γ_ρ, obtaining

$$\frac{k!}{2\pi i} \sum_{n=1}^{\infty} \int_{\gamma_\rho} \frac{f_n(z)}{(z - z_0)^{k+1}}\, dz = \frac{k!}{2\pi i} \int_{\gamma_\rho} \frac{f(z)}{(z - z_0)^{k+1}}\, dz \qquad (k = 0, 1, 2, \ldots). \tag{15.19}$$

For $k = 0$, (15.19) reduces to

$$\frac{1}{2\pi i} \sum_{n=1}^{\infty} \int_{\gamma_\rho} \frac{f_n(z)}{z - z_0}\, dz = \frac{1}{2\pi i} \int_{\gamma_\rho} \frac{f(z)}{z - z_0}\, dz$$

or

$$f(z_0) = \sum_{n=1}^{\infty} f_n(z_0) = \frac{1}{2\pi i} \int_{\gamma_\rho} \frac{f(z)}{z - z_0}\, dz,$$

where we use Cauchy's integral formula (Theorem 14.1) and the fact that every $f_n(z)$ is analytic on $\overline{I(\gamma_\rho)}$. Therefore $f(z)$ is represented by an integral of the Cauchy type (in fact, Cauchy's integral) at every point $z_0 \in G$, and hence, according to Theorem 14.5, $f(z)$ is analytic on G. For $k > 0$, (15.19) reduces to (15.17), with z replaced by z_0, where we now use formula (14.20) and the fact that $f(z)$ itself, as well as every $f_n(z)$, is analytic on $\overline{I(\gamma_\rho)}$.

---

[4] As always, we use the notation

$$f_n^{(0)}(z) = f_n(z), \quad f_n^{(1)}(z) = \frac{df_n(z)}{dz}, \ldots, \quad f_n^{(k)}(z) = \frac{d^k f_n(z)}{dz^k}, \ldots$$

Finally, we must prove that each differentiated series

$$\sum_{n=1}^{\infty} f_n^{(k)}(z) = \frac{k!}{2\pi i} \sum_{n=1}^{\infty} \int_{\gamma_\rho} \frac{f_n(\zeta)}{(\zeta - z)^{k+1}}\, d\zeta = \frac{k!}{2\pi i} \int_{\gamma_\rho} \frac{f(\zeta)}{(\zeta - z)^{k+1}}\, d\zeta \tag{15.20}$$

is uniformly convergent on every compact subset of G. To show this, we shall use Theorem 15.5, choosing the neighborhood $\mathcal{N}(z_0)$ which figures in the statement of the theorem to be the open disk $|z - z_0| < \frac{1}{2}\rho$. Since the original series (15.16) is uniformly convergent on γ_ρ, there is an integer $N(\varepsilon) > 0$ such that

$$|s_n(z) - f(z)| < \varepsilon$$

for all $n > N(\varepsilon)$ and $z \in \gamma_\rho$, where $s_n(z)$ denotes the nth partial sum of (15.16). It follows that

$$\begin{aligned} &\left| \frac{k!}{2\pi i} \sum_{j=1}^{n} \int_{\gamma_\rho} \frac{f_j(\zeta)}{(\zeta - z)^{k+1}}\, d\zeta - \frac{k!}{2\pi i} \int_{\gamma_\rho} \frac{f(\zeta)}{(\zeta - z)^{k+1}}\, d\zeta \right| \\ &\qquad = \left| \frac{k!}{2\pi i} \int_{\gamma_\rho} \frac{s_n(\zeta) - f(\zeta)}{(\zeta - z)^{k+1}}\, d\zeta \right| < \frac{k!}{2\pi} \frac{\varepsilon}{(\rho/2)^{k+1}}\, 2\pi\rho \end{aligned} \tag{15.21}$$

for all $n > N(\varepsilon)$ and $z \in \mathcal{N}(z_0)$. The right-hand side of (15.21) obviously goes to zero with ε itself. Therefore the series (15.20) is uniformly convergent on $\mathcal{N}(z_0)$, and the proof is complete.

Remark. Theorem 15.6 does not hold if G is assumed to be an arbitrary set instead of a *domain*, as shown by the following two examples:

Example 1. Consider the series

$$\sum_{n=0}^{\infty} b^n \cos(a^n \pi x), \tag{15.22}$$

where $0 < b < 1$, a is an odd integer and $ab > 1 + \frac{3}{2}\pi$. Each term of this series is analytic on the real axis (in fact, everywhere in the finite plane, if x is regarded as complex). Moreover, since

$$|b^n \cos(a^n \pi x)| \leqslant b^n,$$

the series (15.22) is uniformly convergent on the real axis, by comparison with the convergent geometric series

$$\sum_{n=0}^{\infty} b^n$$

(see Theorem 15.2). However, as first shown by Weierstrass, the sum $\chi(x)$ of (15.22) is not analytic on the real axis. In fact, $\chi(x)$ does not have a derivative at any point of the real axis.[5]

---

[5] Of course, $\chi(x)$ is continuous on the real axis, since Theorem 15.3 remains valid for functions of a real variable. Thus $\chi(x)$ is a *continuous, nowhere differentiable* function. For further details, see e.g., E. C. Titchmarsh, *The Theory of Functions*, second edition, Oxford University Press, London (1937), Sec. 11.22.

Example 2. Consider the series

$$\sin x + \left[\frac{\sin 2x}{2} - \sin x\right] + \left[\frac{\sin 3x}{3} - \frac{\sin 2x}{2}\right] + \cdots, \qquad (15.23)$$

whose terms are analytic on the real axis (in fact, everywhere in the finite plane if x is regarded as complex). Since

$$\left|\sin x + \left[\frac{\sin 2x}{2} - \sin x\right] + \cdots + \left[\frac{\sin nx}{n} - \frac{\sin (n-1)x}{n-1}\right]\right| = \left|\frac{\sin nx}{n}\right| \leqslant \frac{1}{n}$$

for all real x, the series (15.23) converges uniformly to 0 on the real axis. However, differentiating (15.23) term by term, we obtain the series

$$\cos x + [\cos 2x - \cos x] + \cdots + [\cos nx - \cos (n-1)x] + \cdots, \qquad (15.24)$$

with partial sums

$$\cos x, \cos 2x, \ldots, \cos nx, \ldots$$

If $x \neq 2k\pi$ (where k is an integer), these partial sums form a divergent sequence, while if $x = 2k\pi$, they form a sequence with limit 1. On the other hand, since the sum of (15.23) is identically zero, so is its derivative. Thus term-by-term differentiation of the series (15.23) is not permissible.

Remark. Neither of the series (15.22) and (15.23) can converge uniformly on any domain containing points of the real axis, since otherwise we would arrive at a contradiction with Theorem 15.6.

The following result is often used to establish the uniform convergence of a series on a closed domain:

THEOREM 15.7. *Suppose every term of the series*

$$\sum_{n=1}^{\infty} f_n(z) \qquad (15.25)$$

is analytic on a domain G and continuous on $\bar{G}$. Then the series is uniformly convergent on $\bar{G}$ if it is uniformly convergent on Γ, the boundary of G.

Proof. By hypothesis, given any $\varepsilon > 0$, there is an integer $N(\varepsilon) > 0$ such that

$$|f_{n+1}(z) + \cdots + f_{n+p}(z)| < \varepsilon \qquad (15.26)$$

for all $n > N(\varepsilon)$, $p > 0$ and $z \in \Gamma$. But the function

$$f_{n+1}(z) + \cdots + f_{n+p}(z)$$

is analytic on G and continuous on $\bar{G}$. Therefore, according to Theorem 14.4, Corollary 1, the maximum of the function $|f_{n+1}(z) + \cdots + f_{n+p}(z)|$ on $\bar{G}$ is actually achieved on Γ, and is therefore less than ε, because of (15.26). In other words, (15.26) holds for all $n > N(\varepsilon)$, $p > 0$ and $z \in \bar{G}$, i.e., the series (15.25) is uniformly convergent on $\bar{G}$.

76. Uniformly Convergent Sequences. Improper Integrals of the Cauchy Type

We now turn our attention to (infinite) *sequences* of functions, i.e., to sequences $\{f_n(z)\}$ whose terms are functions of a complex variable defined on a set E in the complex plane. Such a sequence $\{f_n(z)\}$ is said to be *uniformly convergent on* E if, given any $\varepsilon > 0$, there exists an integer $N(\varepsilon) > 0$ such that

$$|f_{n+p}(z) - f_n(z)| < \varepsilon$$

for all $n > N(\varepsilon)$, $p > 0$ and $z \in E$.

Remark. The sequence $\{f_n(z)\}$ can be regarded as the sequence of partial sums corresponding to the *series*

$$f_0(z) + [f_1(z) - f_0(z)] + [f_2(z) - f_1(z)] + \cdots + [f_n(z) - f_{n-1}(z)] + \cdots, \tag{15.27}$$

and in fact, $\{f_n(z)\}$ is a uniformly convergent sequence on E if and only if (15.27) is a uniformly convergent series on E. Moreover, the terms of the sequence $\{f_n(z)\}$ are continuous (or analytic) if and only if the terms of the series (15.27) are continuous (or analytic). It follows that every theorem of Sec. 75 has an obvious counterpart for infinite sequences. In particular, Theorem 15.6 becomes

THEOREM 15.8 (*Weierstrass' theorem on uniformly convergent sequences of analytic functions*). *If the sequence* $\{f_n(z)\}$ *is uniformly convergent on every compact subset of a domain* G, *and if every term* $f_n(z)$ *is analytic on* G, *then the limit function*

$$f(z) = \lim_{n \to \infty} f_n(z)$$

is also analytic on G. *Moreover, as* $n \to \infty$ *each sequence of derivatives* $\{f_n^{(k)}(z)\}$, $k = 1, 2, \ldots$, *converges uniformly to* $f^{(k)}(z)$ *on every compact subset of* G.

THEOREM 15.9. *If the sequence* $\{f_n(z)\}$ *is uniformly convergent on* E, *then so is the sequence* $\{|f_n(z)|\}$. *Moreover, if there exists a constant* $M > 0$ *such that*

$$|\varphi(z)| \leqslant M$$

for all $z \in E$, *then* $\{\varphi(z)f_n(z)\}$ *is uniformly convergent on* E.

Proof. The first assertion follows from the inequality

$$||f_{n+p}(z)| - |f_n(z)|| \leqslant |f_{n+p}(z) - f_n(z)|.$$

The second assertion follows from the fact that if

$$|f_{n+p}(z) - f_n(z)| < \frac{\varepsilon}{M}$$

for all $n > N(\varepsilon)$, $p > 0$ and $z \in E$, then

$$|\varphi(z)f_{n+p}(z) - \varphi(z)f_n(z)| < \frac{\varepsilon}{M} M = \varepsilon$$

for all $n > N(\varepsilon)$, $p > 0$ and $z \in E$.

The sequence $\{f_n(z)\}$ is said to be *uniformly bounded on E* if there exists a constant $M > 0$ such that

$$|f_n(z)| \leqslant M$$

for all n and $z \in E$.

THEOREM 15.10. *If the sequence $\{f_n(z)\}$ is uniformly convergent on a compact set F, and if every term $f_n(z)$ is continuous on F, then $\{f_n(z)\}$ is uniformly bounded on F.*

Proof. The function

$$f(z) = \lim_{n \to \infty} f_n(z)$$

is continuous on F, and hence bounded on F (see Theorem 3.15, Corollary), so that

$$|f(z)| \leqslant M'$$

for all $z \in F$. Moreover, given any $\varepsilon > 0$, there is an integer $N(\varepsilon) > 0$ such that

$$|f_n(z) - f(z)| < \varepsilon$$

for all $n > N(\varepsilon)$ and $z \in F$.[6] Choosing $\varepsilon = 1$ and $\nu = N(1)$, we find that

$$|f_n(z) - f(z)| < 1,$$

i.e.,

$$|f_n(z)| < |f(z)| + 1 < M' + 1,$$

for $n = \nu + 1, \nu + 2, \ldots$ and all $z \in F$. But each of the functions $f_1(z), \ldots, f_\nu(z)$ is separately bounded on F, which means that

$$|f_n(z)| \leqslant M_n$$

for $n = 1, \ldots, \nu$ and all $z \in F$. Therefore,

$$|f_n(z)| \leqslant \max(M_1, \ldots, M_\nu, M' + 1) \equiv M$$

for all n and $z \in F$, and the theorem is proved.

THEOREM 15.11. *If the sequence $\{f_n(z)\}$ is uniformly convergent and uniformly bounded on any set E, then the sequence $\{[f_n(z)]^k\}$, where k is any positive integer, is also uniformly convergent (and uniformly bounded) on E.*

---

[6] Here we use the analogues for sequences of Theorems 15.3 and 15.1.

Proof. We need only note that if $|f_n(z)| \leqslant M$ for all n and $z \in E$, and if

$$|f_{n+p}(z) - f_n(z)| < \frac{\varepsilon}{kM^{k-1}}$$

for all $n > N(\varepsilon)$, $p > 0$ and $z \in E$, then

$$\begin{aligned} |[f_{n+p}(z)]^k - [f_n(z)]^k| &= |\{f_{n+p}(z) - f_n(z)\}\{[f_{n+p}(z)]^{k-1} \\ &\quad + [f_{n+p}(z)]^{k-2}[f_n(z)] + \cdots + [f_n(z)]^{k-1}\}| \\ &\leqslant kM^{k-1}\,|f_{n+p}(z) - f_n(z)| < \varepsilon \end{aligned}$$

for all $n > N(\varepsilon)$, $p > 0$ and $z \in E$ (the uniform boundedness of $\{[f_n(z)]^k\}$ is obvious).

Besides a sequence of functions $\{f_n(z)\}$, where the index n ranges over the set of all positive integers, one can also consider a *family* of functions $\{f_\tau(z)\}$, indexed by a continuously varying parameter τ. Let every function $f_\tau(z)$ be defined on a set E, and suppose, to be explicit, that τ takes any value in the *half-open* interval $[\alpha, \beta)$, i.e., in the interval $\alpha \leqslant \tau < \beta$ (where $\beta \leqslant +\infty$). Then the family $\{f_\tau(z)\}$ is said to be *uniformly convergent on* E as $\tau \to \beta -$,[7] if given any $\varepsilon > 0$, there exists a number $\beta(\varepsilon) < \beta$ such that

$$|f_{\tau'}(z) - f_\tau(z)| < \varepsilon$$

for all $\tau > \beta(\varepsilon)$, $\tau' > \beta(\varepsilon)$ and $z \in E$. It follows that the sequence $\{f_{\tau_n}(z)\}$ is uniformly convergent on E, where $\{\tau_n\}$ is any sequence of points in $[\alpha, \beta)$ converging to β. The limit of the family $\{f_\tau(z)\}$ as $\tau \to \beta -$ is defined as the limit of any such sequence $\{f_{\tau_n}(z)\}$ as $n \to \infty$, and is denoted by

$$\lim_{\tau \to \beta -} f_\tau(z). \tag{15.28}$$

To see that this definition uniquely specifies the limit function (15.28), let $\{\tau_n\}$, $\{\tilde{\tau}_n\}$ be two different sequences of points in $[\alpha, \beta)$ which both converge to β, and suppose that

$$f_{\tau_n}(z) \to f(z), \qquad f_{\tilde{\tau}_n}(z) \to \tilde{f}(z)$$

as $n \to \infty$. Then, given any $\varepsilon > 0$, there exist positive integers N_1, N_2 and N_3 such that the following three inequalities hold for all $z \in E$:

$$|f_{\tau_n}(z) - f(z)| < \frac{\varepsilon}{3} \quad \text{if} \quad n > N_1,$$

$$|f_{\tilde{\tau}_n}(z) - \tilde{f}(z)| < \frac{\varepsilon}{3} \quad \text{if} \quad n > N_2,$$

$$|f_{\tau_n}(z) - f_{\tilde{\tau}_n}(z)| < \frac{\varepsilon}{3} \quad \text{if} \quad n > N_3.$$

---

[7] The minus sign indicates that τ approaches β *from the left*. For generality, we treat the case where τ approaches a point which does not belong to the interval of definition of the family $\{f_\tau(z)\}$. If τ approaches an interior point of $[\alpha, \beta)$, the minus sign is unnecessary.

Choosing $n > \max(N_1, N_2, N_3)$ and combining all these inequalities, we find that

$$|f(z) - \tilde{f}(z)| < \varepsilon$$

for all $z \in E$ and arbitrary ε, i.e., $f(z) \equiv \tilde{f}(z)$ on E, as asserted.

All the results of Sec. 75 can be extended to the case of families of functions, by exploiting this equivalence between uniform convergence of families and uniform convergence of sequences. For example, we have the following analogue of Theorems 15.6 and 15.8:

THEOREM 15.12 (*Weierstrass' theorem on uniformly convergent families of analytic functions*). *If the family* $\{f_\tau(z)\}$, $\alpha \leqslant \tau < \beta$ *is uniformly convergent on every compact subset of a domain* G, *and if every function* $f_\tau(z)$ *is analytic on* G, *then the limit function*

$$f(z) = \lim_{\tau \to \beta -} f_\tau(z)$$

is also analytic on G. *Moreover, as* $\tau \to \beta -$ *each family of derivatives* $\{f_\tau^{(k)}(z)\}$, $k = 1, 2, \ldots$ *converges uniformly to the function* $f^{(k)}(z)$ *on every compact subset of* G.

Example. Consider the curve L with equation

$$z = \lambda(t) \qquad (\alpha \leqslant t < \beta), \tag{15.29}$$

where $\lambda(t)$ is continuous on $[\alpha, \beta)$ and approaches ∞ if and only if $t \to \beta -$; such a curve is said to be *unbounded*. Suppose every *finite arc* of L, i.e., every curve L_τ with equation

$$z = \lambda(t) \qquad (\alpha \leqslant t \leqslant \tau < \beta),$$

is rectifiable. Then, if $\varphi(z)$ is a continuous function on L, the integrals of the Cauchy type

$$f_\tau(z) = \frac{1}{2\pi i} \int_{L_\tau} \frac{\varphi(\zeta)}{\zeta - z}\, d\zeta \qquad (\alpha \leqslant \tau < \beta)$$

define a family of functions which are analytic on any domain G containing no points of L. Suppose that as $\tau \to \beta -$ the family $\{f_\tau(z)\}$ is uniformly convergent on every compact subset $F \subset G$. For example, this will be the case if the improper integral

$$\int_L \varphi(\zeta)\, d\zeta$$

is absolutely convergent, i.e., if the limit

$$\lim_{\tau \to \beta -} \int_{L_\tau} |\varphi(\zeta)|\, ds = \int_L |\varphi(\zeta)|\, ds$$

exists ($ds = |d\zeta|$), since then

$$\begin{aligned} |f_{\tau'}(z) - f_\tau(z)| &= \left| \frac{1}{2\pi i} \int_{-L_\tau + L_{\tau'}} \frac{\varphi(\zeta)}{\zeta - z}\, d\zeta \right| \\ &\leqslant \frac{1}{2\pi\rho} \left| \int_{L_{\tau'}} |\varphi(z)|\, ds - \int_{L_\tau} |\varphi(z)|\, ds \right|, \end{aligned}$$

where ρ is the distance between F and L (cf. Theorem 3.19), and hence we can obviously find a $\beta(\varepsilon) < \beta$ such that

$$|f_{\tau'}(z) - f_{\tau}(z)| < \varepsilon$$

for all $\tau > \beta(\varepsilon)$, $\tau' > \beta(\varepsilon)$ and $z \in F$. Therefore the improper integral

$$f(z) = \lim_{\tau \to \beta -} \frac{1}{2\pi i} \int_{L_\tau} \frac{\varphi(\zeta)}{\zeta - z}\, d\zeta = \frac{1}{2\pi i} \int_{L} \frac{\varphi(\zeta)}{\zeta - z}\, d\zeta$$

exists. Such an integral along an unbounded curve L is called an *improper integral of the Cauchy type*. It follows from Theorem 15.12 that $f(z)$ is analytic on any domain G containing no points of L. Theorem 15.12 also implies that as $\tau \to \beta -$, the family $\{f_\tau^{(k)}(z)\}$ of derivatives

$$f_\tau^{(k)}(z) = \frac{k!}{2\pi i} \int_{L_\tau} \frac{\varphi(\zeta)}{(\zeta - z)^{k+1}}\, d\zeta \qquad (k = 1, 2, \ldots)$$

converges uniformly to $f^{(k)}(z)$ on every compact subset $F \subset G$. On the other hand, uniform convergence of the family $\{f_\tau^{(k)}(z)\}$ means that the improper integral

$$\lim_{\tau \to \beta -} \frac{k!}{2\pi i} \int_{L_\tau} \frac{\varphi(\zeta)}{(\zeta - z)^{k+1}}\, d\zeta$$

exists, and therefore

$$f^{(k)}(z) = \frac{k!}{2\pi i} \int_{L} \frac{\varphi(\zeta)}{(\zeta - z)^{k+1}}\, d\zeta \qquad (k = 1, 2, \ldots).$$

Thus, all the results derived in Sec. 72 for integrals of the Cauchy type remain valid for improper integrals of the Cauchy type.

Remark. These results can be carried over with obvious modifications to the case of a curve L with equation

$$z = \lambda(t) \qquad (\alpha < t < \beta),$$

where $\lambda(t)$ is continuous on (α, β), and approaches ∞ if and only if $t \to \alpha +$ or $t \to \beta -$ (the plus sign indicates that t approaches α *from the right*). In this case, L is an unbounded curve whose initial and final points both lie at infinity (e.g., a straight line or a parabola).

77. Infinite Products

By an *infinite product* we mean an expression of the form

$$\prod_{n=1}^{\infty} u_n, \tag{15.30}$$

with complex nonzero factors u_n (the case of zero factors will be considered at the end of this section). Let

$$\{p_n\} = \{u_1 u_2 \cdots u_n\} \tag{15.31}$$

be the sequence of *partial products* of (15.30). Then the infinite product (15.30) is said to *converge* if the sequence (15.31) converges to a nonzero limit, and the limit of the sequence is called the *value* of the infinite product. Moreover, if

$$\lim_{n \to \infty} p_n = u \neq 0,$$

we write

$$\prod_{n=1}^{\infty} u_n = u.$$

An infinite product which does not converge to a nonzero (finite) limit is said to be *divergent.*

Obviously, a necessary condition for the infinite product (15.31) to converge is that

$$\lim_{n \to \infty} u_n = 1,$$

since if

$$\lim_{n \to \infty} \prod_{k=1}^{n} u_k = u \neq 0,$$

then

$$\lim_{n \to \infty} u_n = \lim_{n \to \infty} \frac{\prod_{k=1}^{n} u_k}{\prod_{k=1}^{n-1} u_k} = 1.$$

In view of this, it is convenient to write the general term of the infinite product (15.31) in the form

$$u_n = 1 + v_n,$$

where now a necessary condition for convergence is that

$$\lim_{n \to \infty} v_n = 0.$$

THEOREM 15.13. *The infinite product*

$$\prod_{n=1}^{\infty} (1 + v_n) \tag{15.32}$$

converges if and only if the infinite series

$$\sum_{n=1}^{\infty} \ln (1 + v_n) \tag{15.33}$$

converges.

Proof. Convergence of (15.33) implies convergence of the two series

$$\sum_{n=1}^{\infty} \ln |1 + v_n|, \qquad \sum_{n=1}^{\infty} \arg (1 + v_n),$$

i.e., of the two sequences $\{s_n\}$, $\{\sigma_n\}$ with terms

$$s_n = \sum_{k=1}^{n} \ln|1 + v_k| = \ln\left|\prod_{k=1}^{n}(1 + v_k)\right|$$

and

$$\sigma_n = \sum_{k=1}^{n} \arg(1 + v_k) = \operatorname{Arg}_n\left[\prod_{k=1}^{n}(1 + v_k)\right], \tag{15.34}$$

where Arg_n denotes that value of the argument for which the equality (15.34) holds. Therefore the sequence

$$\left\{\prod_{k=1}^{n}(1 + v_k)\right\} = \{e^{s_n}(\cos\sigma_n + i\sin\sigma_n)\} \tag{15.35}$$

converges to a nonzero limit, i.e., the infinite product (15.32) converges.

Conversely, if

$$\prod_{n=1}^{\infty}(1 + v_n) = u \neq 0,$$

then the limit

$$\lim_{n\to\infty}|1 + v_1|\cdots|1 + v_n| = |u| \neq 0$$

exists (why?), and hence the series

$$\sum_{n=1}^{\infty}\ln|1 + v_n| \tag{15.36}$$

converges. Moreover, as shown on p. 35, there exists a sequence of values

$$\operatorname{Arg}\left[\prod_{k=1}^{n}(1 + v_k)\right] = \sum_{k=1}^{n}\arg(1 + v_k) + 2\pi\mu_n = \Phi_n$$

(with integral μ_n), converging to any given value Φ of $\operatorname{Arg} u$. Since

$$\begin{aligned}|\Phi_{n+1} - \Phi_n| &= |\arg(1 + v_{n+1}) + 2\pi(\mu_{n+1} - \mu_n)| \\ &\geqslant 2\pi|\mu_{n+1} - \mu_n| - \pi,\end{aligned}$$

and since

$$\lim_{n\to\infty}|\Phi_{n+1} - \Phi_n| = 0,$$

it follows that all the nonnegative integers $|\mu_{n+1} - \mu_n|$ vanish if n exceeds a certain integer $N > 0$, i.e.,

$$\mu_{N+1} = \mu_{N+2} = \cdots = \mu.$$

But then the sequence $\{\Phi_n - 2\pi\mu_n\}$ converges, as well as the sequence $\{\Phi_n\}$ itself, and therefore the series

$$\sum_{n=1}^{\infty}\arg(1 + v_n)$$

converges. This fact, together with the convergence of (15.36), implies the convergence of (15.33).

Remark 1. Suppose the infinite product (15.32) converges. Then, according to Theorem 15.13, the infinite series (15.33) also converges, and hence we can take the limit as $\nu \to \infty$ of the obvious identity

$$\prod_{n=1}^{\nu} (1 + v_n) = \exp\left[\sum_{n=1}^{\nu} \ln (1 + v_n)\right],$$

obtaining

$$\prod_{n=1}^{\infty} (1 + v_n) = \exp\left[\sum_{n=1}^{\infty} \ln (1 + v_n)\right]. \tag{15.37}$$

Remark 2. The infinite product (15.32) is said to be *absolutely convergent* if the infinite series (15.33) is absolutely convergent. Since we can arbitrarily rearrange the terms of an absolutely convergent series without destroying the convergence of the series or the value of its sum, it follows from (15.37) that we can arbitrarily rearrange the factors of an absolutely convergent product without destroying the convergence of the product or changing the value of the product.

THEOREM 15.14. *The infinite product*

$$\prod_{n=1}^{\infty} (1 + v_n)$$

is absolutely convergent if and only if the infinite series

$$\sum_{n=1}^{\infty} v_n \tag{15.38}$$

is absolutely convergent.

Proof. It follows from the expansion

$$\ln (1 + v_n) = v_n - \frac{v_n^2}{2} + \frac{v_n^3}{3} - \cdots = v_n\left(1 - \frac{v_n}{2} + \frac{v_n^2}{3} - \cdots\right),$$

whose validity for complex v_n will be proved in Sec. 81, that

$$\begin{aligned} \frac{1}{2}|v_n| = |v_n|\left(1 - \frac{1}{2^2} - \frac{1}{2^3} - \cdots\right) &< |\ln (1 + v_n)| \\ &< |v_n|\left(1 + \frac{1}{2^2} + \frac{1}{2^3} + \cdots\right) = \frac{3}{2}|v_n| \end{aligned}$$

if $|v_n| < \frac{1}{2}$ (why is this no restriction?). Therefore, as asserted,

$$\sum_{n=1}^{\infty} \ln (1 + v_n)$$

is absolutely convergent if and only if (15.38) is absolutely convergent.

We now turn our attention to infinite products whose factors are functions. An infinite product of the form

$$\prod_{n=1}^{\infty} [1 + v_n(z)], \tag{15.39}$$

where every $v_n(z)$ is a function defined on a given set E, is said to be *uniformly convergent on* E if the sequence of partial products

$$\{p_n(z)\} = \{[1 + v_1(z)] \cdots [1 + v_n(z)]\}$$

is uniformly convergent on E.

THEOREM 15.15. *Suppose every term of the sequence* $\{v_n(z)\}$ *is analytic on a domain* G, *and suppose the infinite series*

$$\sum_{n=1}^{\infty} \ln [1 + v_n(z)] \tag{15.40}$$

is uniformly convergent on every compact subset of G.[8] *Then the infinite product* (15.39) *converges uniformly to a nonvanishing analytic function* $f(z)$ *on every compact subset of* G.

Proof. The convergence of (15.39) to a nonvanishing function $f(z)$ on G follows from Theorem 15.13. The fact that $f(z)$ is analytic on G follows from the analyticity of (15.40) [see Theorem 15.6] and the representation

$$f(z) = \exp\left\{\sum_{n=1}^{\infty} \ln [1 + v_n(z)]\right\}$$

[see (15.37)]. Finally, if F is a compact subset of G, let

$$M = \max_{z \in F} |f(z)|$$

(see Theorem 3.16), and let ε be an arbitrary positive number less than M. Since (15.40) is uniformly convergent on F, there is an integer $N(\varepsilon) > 0$ such that

$$\left| \sum_{k=n+1}^{\infty} \ln [1 + v_k(z)] \right| < \frac{\varepsilon}{2M} < \frac{1}{2}$$

for all $n > N(\varepsilon)$ and $z \in F$. Therefore, anticipating the validity of the expansion

$$e^z = 1 + z + \frac{1}{2!} z^2 + \frac{1}{3!} z^3 + \cdots$$

[8] In particular, none of the terms $v_n(z)$ can take the value -1 at any point of G.

for complex z (see Sec. 81), we find that

$$\left|f(z) - \prod_{k=1}^{n} [1 + v_k(z)]\right| = \left|\exp\left\{\sum_{k=1}^{\infty} \ln [1 + v_k(z)]\right\} - \exp\left\{\sum_{k=1}^{n} \ln [1 + v_k(z)]\right\}\right|$$

$$= \left|\exp\left\{\sum_{k=1}^{\infty} \ln [1 + v_k(z)]\right\}\right| \left|1 - \exp\left\{-\sum_{k=n+1}^{\infty} \ln [1 + v_k(z)]\right\}\right|$$

$$\leqslant M\left[\frac{\varepsilon}{2M} + \frac{1}{2!}\left(\frac{\varepsilon}{2M}\right)^2 + \frac{1}{3!}\left(\frac{\varepsilon}{2M}\right)^3 + \cdots\right]$$

$$= \frac{\varepsilon}{2}\left[1 + \frac{1}{2!}\frac{\varepsilon}{2M} + \frac{1}{3!}\left(\frac{\varepsilon}{2M}\right)^2 + \cdots\right] < \frac{\varepsilon}{2}\left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots\right) = \varepsilon$$

for all $n > N(\varepsilon)$ and $z \in F$, i.e., the infinite product (15.39) is uniformly convergent on F.

THEOREM 15.16. *Suppose every term of the sequence $\{v_n(z)\}$ is analytic on a domain G, and suppose none of the terms $v_n(z)$ takes the value -1 at any point of G. Moreover, suppose there is a convergent series*

$$\sum_{n=1}^{\infty} M_n,$$

whose terms are nonnegative constants, such that

$$|v_n(z)| \leqslant M_n$$

for all $z \in G$ and all n exceeding a certain integer $N > 0$. Then the infinite product

$$\prod_{n=1}^{\infty} [1 + v_n(z)]$$

converges uniformly on G to a nonvanishing analytic function $f(z)$.

Proof. Since

$$|\ln [1 + v_n(z)]| < \tfrac{3}{2}|v_n(z)| < \tfrac{3}{2}M_n \qquad (z \in G)$$

if $|v_n(z)| < \frac{1}{2}$ (see the proof of Theorem 15.14), and since $|v_n(z)| < \frac{1}{2}$ for all $z \in G$ if n is sufficiently large, the infinite series

$$\sum_{n=1}^{\infty} \ln [1 + v_n(z)]$$

is uniformly convergent on G, by Weierstrass' M-test (Theorem 15.2). It follows from Theorem 15.15 that the infinite product

$$\prod_{n=1}^{\infty} [1 + v_n(z)] \tag{15.41}$$

converges uniformly on G to a nonvanishing analytic function, as asserted. We note, in passing, that (15.41) is also absolutely convergent on G.

Finally, we enlarge the class of infinite products by allowing a *finite* number of zero factors. Given an infinite product

$$\prod_{n=1}^{\infty} u_n \tag{15.42}$$

with finitely many zero factors, let $N \geqslant 0$ be any integer such that all the factors $u_{N+1}, u_{N+2}, \ldots$ are nonzero, and let ν be any integer greater than N. Then we say that (15.42) *converges* if the limit[9]

$$\lim_{\nu \to \infty} \prod_{n=1}^{\nu} u_n = \lim_{\nu \to \infty} \prod_{n=1}^{N} u_n \cdot \prod_{n=N+1}^{\nu} u_n = \prod_{n=1}^{N} u_n \cdot \lim_{\nu \to \infty} \prod_{n=N+1}^{\nu} u_n \tag{15.43}$$

exists, i.e., if the infinite product

$$\prod_{n=N+1}^{\infty} u_n \tag{15.44}$$

converges in the sense of our previous definition. Naturally, if (15.44) converges, it follows from (15.43) that

$$\prod_{n=1}^{\infty} u_n = \lim_{\nu \to \infty} \prod_{n=1}^{\nu} u_n = 0,$$

unless $N = 0$. With this extended definition of convergence of an infinite product, we can assert that *an infinite product vanishes if and only if at least one of its factors vanishes.*

PROBLEMS

15.1. Find the set on which each of the following series converges:

a) $\sum_{n=1}^{\infty} \frac{z^n}{1 - z^n}$; b) $\sum_{n=1}^{\infty} \frac{\sin nz}{n}$; c) $\sum_{n=1}^{\infty} \frac{\cos nz}{n^2}$; d) $\sum_{n=1}^{\infty} \frac{(-1)^n}{z + n}$; e) $\sum_{n=1}^{\infty} \frac{2^n}{z^{2^n} + 1}$.

Hint. In part e) use the fact that

$$\frac{2^n}{z^{2^n} - 1} = \frac{2}{z^2 - 1} - \frac{2^2}{z^2 + 1} - \cdots - \frac{2^{n-1}}{z^{2^{n-1}} + 1}.$$

15.2. Where does each of the series given in the preceding problems converge *uniformly*?

15.3. Prove that if the series

$$\sum_{n=1}^{\infty} f_n(z) = f(z)$$

---

[9] To cover the case where *no* u_n is zero, the otherwise meaningless expression

$$\prod_{n=1}^{0} u_n$$

is *defined* as equal to 1 (for any choice of u_n).

is uniformly convergent on a set E, and if $|\varphi(z)| \leqslant \text{const}$ on E, then the series

$$\sum_{n=1}^{\infty} \varphi(z) f_n(z)$$

is uniformly convergent on E, with sum $\varphi(z) f(z)$.

Comment. Cf. Theorem 15.9.

15.4. Prove that uniform convergence of the series

$$\sum_{n=1}^{\infty} |f_n(z)|$$

on a set E implies uniform convergence of the series

$$\sum_{n=1}^{\infty} f_n(z)$$

on E, but not conversely.

Hint. Consider the series

$$\sum_{n=1}^{\infty} \frac{z^n}{n}$$

in the interval $(-1, 0)$.

15.5. Prove that if the series

$$\sum_{n=1}^{\infty} |f_n(z)|$$

is uniformly convergent on every compact subset of a domain G, and if every term $f_n(z)$ is analytic on G, then each series

$$\sum_{n=1}^{\infty} |f_n^{(k)}(z)| \qquad (k = 1, 2, \ldots)$$

is uniformly convergent on every compact subset of G.

15.6. Suppose the series

$$\sum_{n=1}^{\infty} f_n(z) = f(z) \tag{15.45}$$

is uniformly convergent on every compact subset of a *simply connected* domain G, and suppose every term $f_n(z)$ is analytic on G. Prove that *indefinite* term-by-term integration of (15.45) is permissible, i.e., prove that

$$\sum_{n=1}^{\infty} \int_{z_0}^{z} f_n(\zeta)\, d\zeta = \int_{z_0}^{z} f(\zeta)\, d\zeta, \tag{15.46}$$

where z_0 is a fixed point in G, and z a variable point in G. Moreover, prove that the series in the left-hand side of (15.46) converges uniformly to the function in the right-hand side on every compact subset of G.

Hint. Let z_1 be another fixed but arbitrary point of G, and choose $\rho > 0$ such that G contains the circle $\gamma_\rho : |z - z_1| = \rho$ and its interior $I(\gamma_\rho)$. Integrate from z_0 to z_1 along any fixed rectifiable curve $L \subset G$, and then along the line segment joining z_1 to z. Finally use Theorem 15.5.

17.7. Generalize the result of the preceding problem to the case where G is multiply connected.

Hint. We must now consider the possibility that the integrals in (15.46) define multiple-valued functions. However, the same proof is valid if the variable point z is confined to a simply connected subdomain $D \subset G$ (cf. Theorem 13.7). In fact, we need only integrate along a fixed path L^* joining z_0 to an arbitrary point $z_0^* \in D$, thereby specifying a single-valued branch on D, and then argue as before, with z_0^* playing the role of z_0 and D that of G.

15.8. Where does each of the following sequences converge uniformly:

a) $\left\{\frac{1}{1+z^n}\right\}$; b) $\left\{\frac{z^n}{1+z^{2n}}\right\}$; c) $\left\{\frac{\sin nz}{n}\right\}$?

15.9. Why is there no contradiction between Theorem 15.9 and the result of Problem 15.4?

15.10. Suppose the improper integral of the Cauchy type

$$f(z) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{\varphi(x)}{x - z}\,dx, \qquad \varphi(x) = \frac{x^2}{x^2 + 1}$$

is defined as the *Cauchy principal value*[10]

$$f(z) = \lim_{R \to \infty} \frac{1}{2\pi i}\int_{-R}^{R} \frac{\varphi(x)}{x - z}\,dx.$$

Prove that

$$f(z) = \begin{cases} \dfrac{z}{2(z+i)} & \text{if } \operatorname{Im} z > 0, \\ -\dfrac{z}{2(z-i)} & \text{if } \operatorname{Im} z < 0. \end{cases}$$

Hint. Write

$$\int_{-R}^{R} \frac{\varphi(x)}{x - z}\,dx = \int_{-R}^{R} \frac{\varphi(x) - \varphi(\infty)}{x - z}\,dx + \varphi(\infty)\int_{-R}^{R} \frac{dx}{x - z},$$

and take the limit $R \to \infty$. Use Cauchy's integral theorem to evaluate the first integral on the right.

15.11. Evaluate the following infinite products:

a) $\prod_{n=2}^{\infty}\left(1 - \frac{1}{n^2}\right)$; b) $\prod_{n=1}^{\infty}\left(1 + \frac{1}{n(n+2)}\right)$; c) $\prod_{n=3}^{\infty}\frac{n^2 - 4}{n^2 - 1}$;

d) $\prod_{n=2}^{\infty}\left(1 - \frac{2}{n(n+1)}\right)$; e) $\prod_{n=2}^{\infty}\frac{n^3 - 1}{n^3 + 1}$; f) $\prod_{n=1}^{\infty}\left(1 + \frac{(-1)^{n+1}}{n}\right)$.

Ans. b) 2; d) $\frac{1}{3}$; f) 1.

15.12. Give an example showing that

$$\lim_{n \to \infty} u_n = 1$$

[10] Note that the considerations given in the example on p. 333 do not apply here, since $\int_{-\infty}^{\infty} \varphi(x)\,dx$ diverges.

is not a sufficient condition for the infinite product

$$\prod_{n=1}^{\infty} u_n \tag{15.47}$$

to converge.

15.13. Prove that the infinite product (15.47) converges if and only if given any $\varepsilon > 0$, there exists an integer $N(\varepsilon) > 0$ such that

$$|u_{n+1}u_{n+2} \cdots u_{n+p} - 1| < \varepsilon$$

for all $n > N(\varepsilon)$ and $p > 0$.

Hint. Cf. Theorem 3.4.

15.14. Prove that the infinite product

$$\prod_{n=1}^{\infty} (1 + v_n) \tag{15.48}$$

is absolutely convergent (as defined on p. 337) if and only if the infinite product

$$\prod_{n=1}^{\infty} (1 + |v_n|) \tag{15.49}$$

is convergent.

Hint. Using the inequality $1 + x \leqslant e^x$ for $x \geqslant 0$, prove that (15.49) is convergent if and only if

$$\sum_{n=1}^{\infty} |v_n|$$

is convergent. Then use Theorem 15.14.

Comment. Thus, we could just as well have defined an absolutely convergent infinite product (15.48) as one for which (15.49) converges.

15.15. Where does each of the following infinite products converge absolutely:

a) $\prod_{n=1}^{\infty} (1 - z^n)$; b) $\prod_{n=0}^{\infty} (1 + c_n z)$, where $\prod_{n=0}^{\infty} |c_n|$ converges;

c) $\prod_{n=1}^{\infty} \cos \frac{z}{n}$; d) $\prod_{n=1}^{\infty} \frac{\sin (z/n)}{z/n}$; e) $\prod_{n=1}^{\infty} \left(1 - \frac{1}{n^z}\right)$?

Comment. Here we allow a finite number of zero factors (as at the end of Sec. 77).

15.16. Prove that the infinite product

$$\prod_{n=1}^{\infty} (1 + z^{2^n})$$

is uniformly convergent on every compact subset of the unit disk $K: |z| < 1$, and hence represents an analytic function on K. What is this function?

Ans. $\dfrac{1}{1 - z}$.

16

POWER SERIES: RUDIMENTS

78. The Cauchy-Hadamard Formula

In the preceding chapter, we considered series whose terms are analytic functions. Of all such series, the simplest, and at the same time, the most important, are the power series. By a *power series* we mean a series of the form

$$a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots, \tag{16.1}$$

where $a_0, a_1, \ldots, a_n, \ldots$ and z_0 are given complex numbers.

THEOREM 16.1 (*Cauchy-Hadamard theorem*). *Given the power series* (16.1), *let*

$$R = \frac{1}{\Lambda}, \quad \textit{where} \quad \Lambda = \overline{\lim_{n \to \infty}} \sqrt[n]{|a_n|}, \tag{16.2}$$

and let γ *be the circle* $|z - z_0| = R$, *with interior* $I(\gamma)$ *and exterior* $E(\gamma)$. *Then there are three possibilities*:

1. *If* $R = 0$, *the series is divergent for all* $z \neq z_0$;
2. *If* $0 < R < \infty$, *the series is absolutely convergent for all* $z \in I(\gamma)$ *and divergent for all* $z \in E(\gamma)$;
3. *If* $R = \infty$, *the series is absolutely convergent for all finite* z.

Proof. We examine each of the three possibilities in turn (obviously, in every case, the series is absolutely convergent for $z = z_0$):

1. $R = 0$, $\Lambda = \infty$. In this case,

$$\Lambda > \frac{1}{|z - z_0|}$$

for any $z \neq z_0$, i.e., for any $z \neq z_0$ we can find an infinite sequence $\{n_k\}$ of positive integers, where $n_1 < \cdots < n_k < \cdots$, such that

$$\sqrt[n_k]{|a_{n_k}|} > \frac{1}{|z - z_0|} \qquad (k = 1, 2, \ldots). \tag{16.3}$$

But this implies

$$|a_{n_k}(z - z_0)^{n_k}| > 1 \qquad (k = 1, 2, \ldots), \tag{16.4}$$

and hence the necessary condition

$$\lim_{n \to \infty} a_n(z - z_0)^n = 0 \tag{16.5}$$

for the convergence of the series (16.1) cannot be satisfied for any $z \neq z_0$. Therefore, if $R = 0$, the series diverges for all $z \neq z_0$, and converges only at the single point $z = z_0$.

2. $0 < R < \infty$, $0 < \Lambda < \infty$. In this case, if $z \in I(\gamma)$, we can write

$$|z - z_0| = \frac{\theta^2}{\Lambda},$$

where $0 < \theta < 1$ (except for the trivial case $z = z_0$). Then

$$\Lambda < \frac{\Lambda}{\theta} = \frac{\theta}{|z - z_0|},$$

and hence all the values of $\sqrt[n]{|a_n|}$ starting from some value of n must be less than

$$\frac{\theta}{|z - z_0|},$$

i.e., the inequality

$$|a_n(z - z_0)^n| < \theta^n$$

holds, starting from some value of n. Therefore, since the geometric series

$$1 + \theta + \cdots + \theta^n + \cdots$$

is convergent, the series (16.1) is absolutely convergent if $z \in I(\gamma)$. On the other hand, if $z \in E(\gamma)$, then

$$\Lambda > \frac{1}{|z - z_0|},$$

and just as in the case $R = 0$, we can find an infinite sequence $\{n_k\}$ of positive integers, where $n_1 < \cdots < n_k < \cdots$, such that (16.3) and (16.4) hold. Therefore the necessary condition (16.5) for convergence of the series cannot be satisfied, and the series diverges.

3. $R = \infty$, $\Lambda = 0$. In this case,

$$\Lambda < \frac{\theta}{|z - z_0|}$$

for any $z \neq z_0$ and any θ $(0 < \theta < 1)$. Therefore the series (16.1) converges absolutely for any $z \neq z_0$ by exactly the same argument as when $0 < R < \infty$, $z \in I(\gamma)$.

Remark 1. The circle γ is called the *circle of convergence* and R is called the *radius of convergence* [of the series (16.1)]. The formula (16.2) is called the *Cauchy-Hadamard formula*, and expresses the radius of convergence in terms of the coefficients a_n of the power series. Obviously, the behavior of the series for $R = 0$ and $R = \infty$ can be regarded as the appropriate limiting case of its behavior for $0 < R < \infty$. It should be noted that Theorem 16.1 says nothing about the behavior of the series on its circle of convergence, i.e., for $z \in \gamma$.

Remark 2. In applying the Cauchy-Hadamard formula, the following relation often turns out to be useful:

$$\lim_{n \to \infty} \sqrt[n]{\frac{n!}{n^n}} = \frac{1}{e}. \tag{16.6}$$

To prove (16.6), we first note that

$$e^n = 1 + \frac{n}{1!} + \cdots + \frac{n^{n-1}}{(n-1)!} + \frac{n^n}{n!}\left[1 + \frac{n}{n+1} + \frac{n^2}{(n+1)(n+2)} + \cdots\right],$$

which implies

$$e^n < n\frac{n^n}{n!} + \frac{n^n}{n!}\left[1 + \frac{n}{n+1} + \left(\frac{n}{n+1}\right)^2 + \cdots\right] = (2n+1)\frac{n^n}{n!}.$$

On the other hand, it is obvious that

$$e^n > \frac{n^n}{n!},$$

and hence

$$\frac{1}{e^n} < \frac{n!}{n^n} < \frac{2n+1}{e^n},$$

from which (16.6) follows immediately.[1]

Example 1. The radius of convergence of the series

$$z + \frac{2^2}{2!}z^2 + \cdots + \frac{n^n}{n!}z^n + \cdots \tag{16.7}$$

---

[1] Actually, it can be shown that

$$\frac{n!}{n^n} = \frac{\sqrt{2\pi n}}{e^n}(1 + \varepsilon),$$

where $\varepsilon \to 0$ as $n \to \infty$, a result known as *Stirling's formula.* See e.g., D. V. Widder, *op. cit.*, p. 386.

equals $1/e$, as we see at once from (16.2) and (16.6). This can also be shown by using the ratio test (see p. 38), since the ratio of the $(n+1)$th term of the series to the nth term has absolute value

$$\left(1+\frac{1}{n}\right)^n |z|,$$

which approaches $e|z|$ as $n \to \infty$. Therefore the series (16.7) is absolutely convergent if $|z| < 1/e$ and divergent if $|z| > 1/e$, i.e., the radius of convergence of the series is $1/e$.

Example 2. Consider the series

$$z + \frac{z^{2^2}}{2!} + \cdots + \frac{z^{n^2}}{n!} + \cdots. \tag{16.8}$$

In this case,

$$\overline{\lim_{n\to\infty}} \sqrt[n]{|a_n|} = \lim_{n\to\infty} \sqrt[n^2]{\frac{1}{n!}} = \lim_{n\to\infty} \sqrt[n^2]{\frac{n^n}{n!}\frac{1}{n^n}} = 1,$$

and the radius of convergence equals 1. The same result can be obtained by using the ratio test, since the ratio of the $(n+1)$th term of the series (16.8) to the nth term has absolute value

$$\frac{|z|^{2n+1}}{n+1},$$

which, as $n \to \infty$, approaches 0 if $|z| < 1$ and ∞ if $|z| > 1$.

From now on, we shall only consider power series for which $R > 0$, i.e., $\Lambda < \infty$. Then the Cauchy-Hadamard formula

$$\overline{\lim_{n\to\infty}} \sqrt[n]{|a_n|} = \frac{1}{R}$$

implies that given any $\varepsilon > 0$ ($\varepsilon < R$), there exists an integer $N(\varepsilon) > 0$ such that

$$|a_n| < \frac{1}{(R-\varepsilon)^n}$$

for all $n > N(\varepsilon)$ if $R < \infty$, or such that

$$|a_n| < \varepsilon^n$$

for all $n > N(\varepsilon)$ if $R = \infty$. In other words, the coefficients of a power series with a nonzero radius of convergence R grow no faster than the terms of a geometric progression with ratio $1/(R-\varepsilon)$ if $R < \infty$ or ratio ε if $R = \infty$, where ε is an arbitrarily small positive number.

THEOREM 16.2. *Suppose the power series* (16.1) *is convergent for* $z = z_1 \neq z_0$, *and let* γ^* *be the circle* $|z - z_0| = |z_1 - z_0|$. *Then the series is absolutely convergent on* $I(\gamma^*)$.

Proof. Let $\gamma: |z - z_0| = R$ be the circle of convergence of the series (16.1). Then $R > 0$, since $z_1 \neq z_0$. According to Theorem 16.1, either $z_1 \in I(\gamma)$, i.e., $\gamma^* \subset I(\gamma)$, or $z_1 \in \gamma$, i.e., $\gamma^* = \gamma$, since the series diverges if $z_1 \in E(\gamma)$. In either case $I(\gamma^*) \subset I(\gamma)$, and hence the series is absolutely convergent on $I(\gamma^*)$.

THEOREM 16.3. *Let $\gamma: |z - z_0| = R$ be the circle of convergence of the power series* (16.1). *Then the series is uniformly convergent on every compact subset of $I(\gamma)$.*

Proof. Since any compact set $F \subset I(\gamma)$ is contained in some closed disk $|z - z_0| \leqslant r < R$, if r is near enough to R, we need only prove that the series is uniformly convergent on every such closed disk. Thus, given a closed disk $|z - z_0| \leqslant r < R$, let ζ be a point such that

$$r < |\zeta - z_0| = \rho < R.$$

Since the series (16.1) is absolutely convergent for $z = \zeta$, the series

$$\sum_{n=0}^{\infty} |a_n| \, |\zeta - z_0|^n = \sum_{n=0}^{\infty} |a_n| \rho^n$$

converges. Therefore, by Weierstrass' M-test (Theorem 15.2), the series (16.1) is uniformly convergent on the disk $|z - z_0| \leqslant r$, since

$$|a_n(z - z_0)^n| < |a_n| \rho^n$$

for all z in the disk $|z - z_0| \leqslant r$.

Remark. The power series (16.1) need not converge uniformly on $I(\gamma)$ itself. In fact, as we already know (see p. 326), the geometric series

$$\sum_{n=0}^{\infty} z^n,$$

which is a power series with radius of convergence 1, is not uniformly convergent on the unit disk $K: |z| < 1$, but only on every compact subset of K.

79. Taylor's Series. The Uniqueness Property

Since the power series

$$f(z) = a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots, \tag{16.9}$$

with radius of convergence R given by formula (16.2), is a series of analytic functions (in fact, entire functions) which converges uniformly on every compact subset of the disk $K: |z - z_0| < R$, it follows from Theorem 15.6 that the sum function $f(z)$ is analytic on K, and that $f^{(k)}(z)$ can be calculated

by differentiating (16.9) term by term, i.e.,

$$f^{(k)}(z) = k!\,a_k + \frac{(k+1)!a_{k+1}}{1!}(z - z_0) + \cdots + \frac{n!\,a_n}{(n-k)!}(z - z_0)^{n-k} + \cdots \tag{16.10}$$

[cf. (1.3)], where the series (16.10) is uniformly convergent on every compact subset of K. Setting $z = z_0$ in (16.10), we obtain

$$f^{(k)}(z_0) = k!\,a_k,$$

and hence

$$a_k = \frac{f^{(k)}(z_0)}{k!} \qquad (k = 0, 1, 2, \ldots). \tag{16.11}$$

[The fact that (16.11) holds for $k = 0$ can be seen by setting $z = z_0$ in (16.9) itself.] Substituting (16.11) in (16.9), we find that

$$f(z) = f(z_0) + \frac{f'(z_0)}{1!}(z - z_0) + \cdots + \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n + \cdots,$$

where the series on the right is known as the *Taylor series* (*expansion*) of the function $f(z)$, at the point z_0. Thus we have proved that every power series is the Taylor series of its sum $f(z)$. These results are summarized in

THEOREM 16.4. *The power series* (16.9), *with radius of convergence* R, *is an analytic function* $f(z)$ *on the disk* $K: |z - z_0| < R$, *and the coefficients* $a_0, a_1, \ldots, a_n, \ldots$ *are given by the formula* (16.11). *Moreover, the series* (16.9) *can be differentiated term by term any number of times, and each differentiated series* (16.10) *is uniformly convergent on every compact subset of* K.

Now suppose two power series

$$a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots \tag{16.12}$$

and

$$b_0 + b_1(z - z_0) + \cdots + b_n(z - z_0)^n + \cdots, \tag{16.13}$$

with radii of convergence R_1 and R_2, respectively, have the same sum in a neighborhood of the point z_0, i.e., suppose

$$a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots = b_0 + b_1(z - z_0) + \cdots + b_n(z - z_0)^n + \cdots$$

if $|z - z_0| < r$, where $0 < r \leqslant \min(R_1, R_2)$. Then, according to (16.11),

$$a_0 = b_0 = f(z_0), \quad a_1 = b_1 = \frac{f'(z_0)}{1!}, \ldots, \quad a_n = b_n = \frac{f^{(n)}(z_0)}{n!}, \ldots,$$

where $f(z)$ denotes the common sum of the two series. Therefore coefficients of the two series with identical indices are identical, and it follows that the two series have the same radius of convergence $R_1 = R_2$. (Note that it is

vital that the point z_0 be the same for both series.) Thus we have proved the following theorem, typical of a class of results known as *uniqueness theorems*:

THEOREM 16.5. *If the sums of two power series in the variable $z - z_0$ coincide in a neighborhood of the point z_0, then identical powers of $z - z_0$ have identical coefficients, i.e., there is a unique power series in the variable $z - z_0$ which has a given sum in a neighborhood of z_0.*

Example. Given a power series

$$f(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n + \cdots, \tag{16.14}$$

suppose $f(z)$ is an *even* function, which means that $f(z) \equiv f(-z)$. If we replace z by $-z$, (16.14) becomes

$$f(-z) = a_0 - a_1 z + a_2 z^2 + \cdots + (-1)^n a_n z^n + \cdots. \tag{16.15}$$

Equating (16.14) and (16.15), we find that

$$a_n = (-1)^n a_n \qquad (n = 0, 1, 2, \ldots),$$

so that

$$a_{2k+1} = -a_{2k+1} \quad \text{or} \quad a_{2k+1} = 0,$$

i.e., coefficients of odd powers of z vanish, and (16.14) reduces to

$$f(z) = a_0 + a_2 z^2 + a_4 z^4 + \cdots + a_{2k} z^{2k} + \cdots.$$

Similarly, if $f(z)$ is an *odd* function, which means that $f(z) \equiv -f(-z)$, we find that coefficients of even powers of z vanish, so that (16.14) reduces to

$$f(z) = a_1 z + a_3 z^3 + \cdots + a_{2k+1} z^{2k+1} + \cdots.$$

We now give a more general formulation of the uniqueness property expressed by Theorem 16.5: Series expansions of the form

$$a_0 \varphi_0(z) + a_1 \varphi_1(z) + \cdots + a_n \varphi_n(z) + \cdots, \tag{16.16}$$

with respect to given functions

$$\varphi_0(z), \varphi_1(z), \ldots, \varphi_n(z), \ldots, \tag{16.17}$$

are said to have the *uniqueness property* on a set E if, whenever the sums of two such series (16.16) and

$$b_0 \varphi_0(z) + b_1 \varphi_1(z) + \cdots + b_n \varphi_n(z) + \cdots$$

coincide for all $z \in E$, the corresponding coefficients coincide, i.e.,

$$a_0 = b_0, \quad a_1 = b_1, \ldots, \quad a_n = b_n, \ldots$$

Writing the relation

$$a_0 \varphi_0(z) + a_1 \varphi_1(z) + \cdots + a_n \varphi_n(z) + \cdots = b_0 \varphi_0(z) + b_1 \varphi_1(z) + \cdots + b_n \varphi_n(z) + \cdots$$

in the form

$$c_0\varphi_0(z) + c_1\varphi_1(z) + \cdots + c_n\varphi_n(z) + \cdots = 0, \tag{16.18}$$

where

$$c_0 = a_0 - b_0, \quad c_1 = a_1 - b_1, \ldots, \quad c_n = a_n - b_n, \ldots,$$

we can define the uniqueness property somewhat differently: Series expansions of the form (16.16) with respect to given functions (16.17) are said to have the uniqueness property on a set E if, whenever the sum of a series

$$c_0\varphi_0(z) + c_1\varphi_1(z) + \cdots + c_n\varphi_n(z) + \cdots$$

vanishes, the coefficients of the series also vanish, i.e.,

$$c_0 = c_1 = \cdots = c_n = \cdots = 0.$$

In this language, Theorem 16.5 states that series expansions with respect to the functions

$$\varphi_n(z) = (z - z_0)^n \qquad (n = 0, 1, 2, \ldots)$$

have the uniqueness property on every disk with center z_0.

Example. To show that Theorem 16.5 is by no means self-evident, we now exhibit a case where the uniqueness property does not hold. Let the functions (16.17) be

$$\varphi_0(z) = 1, \quad \varphi_1(z) = z - 1, \quad \varphi_2(z) = z^2 - z, \ldots, \quad \varphi_n(z) = z^n - z^{n-1}, \ldots, \tag{16.19}$$

let E be the unit disk $|z| < 1$, and consider the series

$$\varphi_0(z) + \varphi_1(z) + \cdots + \varphi_n(z) + \cdots, \tag{16.20}$$

which is of the form (16.16) with

$$a_0 = 1, \quad a_1 = 1, \ldots, \quad a_n = 1, \ldots$$

The nth partial sum of (16.20), i.e.,

$$\varphi_0(z) + \varphi_1(z) + \varphi_2(z) + \cdots + \varphi_n(z) = 1 + (z - 1) + (z^2 - z) + \cdots + (z^n - z^{n-1}),$$

equals z^n, and therefore converges to zero on E (in fact, uniformly on every compact subset of E). It follows that if the series (16.16) with respect to the functions (16.19) converges on E to a sum $f(z)$, then the series

$$(a_0 + \lambda)\varphi_0(z) + (a_1 + \lambda)\varphi_1(z) + \cdots + (a_n + \lambda)\varphi_n(z) + \cdots,$$

where λ is an arbitrary complex number, also converges on E to the same sum $f(z)$. In other words, expansions with respect to the functions (16.19) do not have the uniqueness property, since two series can have the same sum on E without any coefficient of one series being equal to the corresponding coefficient of the other series.

Remark. Suppose series expansions with respect to a given set of functions $\varphi_0(z), \varphi_1(z), \ldots, \varphi_n(z), \ldots$ have the uniqueness property (on some set E), and suppose the coefficients $a_0, a_1, \ldots, a_n, \ldots$ figuring in a convergent series

$$a_0\varphi_0(z) + a_1\varphi_1(z) + \cdots + a_n\varphi_n(z) + \cdots \tag{16.21}$$

are unknown. Then these coefficients can often be found by using the *method of undetermined coefficients.* In this method, various operations are performed on the series (16.21) and on other series with known coefficients, until a relation of the form (16.18) is obtained, where the coefficients $c_0, c_1, \ldots, c_n, \ldots$ are certain combinations of the unknowns $a_0, a_1, \ldots, a_n, \ldots$ (the "undetermined coefficients") and other known numbers. Because of the uniqueness property, (16.18) then implies the infinite system of equations

$$c_0 = 0, \quad c_1 = 0, \ldots, \quad c_n = 0, \ldots,$$

which can often be solved for the coefficients $a_0, a_1, \ldots, a_n, \ldots$.

In the simplest case, where the power series (16.12) and (16.13) reduce to polynomials

$$a_0 + a_1(z - z_0) + \cdots + a_m(z - z_0)^m \qquad (a_m \neq 0)$$

and

$$b_0 + b_1(z - z_0) + \cdots + b_n(z - z_0)^n \qquad (b_n \neq 0),$$

the requirement that the values of these two polynomials coincide for all z belonging to a certain neighborhood of z_0, as assumed in Theorem 16.5, is unnecessary. In fact, if $N = \max(m, n)$, and if the values of the two polynomials are the same for more than N different values of z, then $m = n$ and the polynomials are identically equal, i.e.,

$$a_0 = b_0, \quad a_1 = b_1, \ldots, \quad a_n = b_n$$

(cf. Problem 9.3). In the case of power series, which in many respects can be regarded as "polynomials of infinitely high degree," one can deduce that two power series have identical coefficients only if it is known that the two series have the same sum on an infinite set of points. However, even in this case, it is not necessary to assume that the sums of the two series coincide at all points of some neighborhood of z_0. Actually, as shown by the following generalization of Theorem 16.5, it is only necessary that the sums of the series (16.12) and (16.13) coincide on a set of points which has z_0 as a limit point, for example, on the set of points

$$z_n = z_0 + \frac{1}{n} \qquad (n = 1, 2, \ldots).$$

THEOREM 16.6. *If the sums of two power series in the variable $z - z_0$ coincide on a set of points E for which z_0 is a limit point, then identical powers of $z - z_0$ have identical coefficients, i.e., there is a unique power series in the variable $z - z_0$ which has a given sum on the set E.*

Proof. Let the two power series be of the form

$$a_0 + a_1(z - z_0) + a_2(z - z_0)^2 + \cdots,$$
$$b_0 + b_1(z - z_0) + b_2(z - z_0)^2 + \cdots,$$

and suppose that

$$\lim_{n \to \infty} z_n = z_0,$$

where $z_n \in E$, $z_n \neq z_0$ $(n = 1, 2, \ldots)$. Then, since

$$\begin{aligned} a_0 + a_1(z_n - z_0) + a_2(z_n - z_0)^2 + \cdots \\ = b_0 + b_1(z_n - z_0) + b_2(z_n - z_0)^2 + \cdots \end{aligned} \tag{16.22}$$

for all $n = 1, 2, \ldots$, and since the sum of a power series is continuous on the interior of its circle of convergence (cf. Theorem 15.3), we have

$$\begin{aligned} a_0 &= \lim_{n \to \infty} [a_0 + a_1(z_n - z_0) + a_2(z_n - z_0)^2 + \cdots] \\ &= \lim_{n \to \infty} [b_0 + b_1(z_n - z_0) + b_2(z_n - z_0)^2 + \cdots] = b_0, \end{aligned}$$

i.e.,

$$a_0 = b_0.$$

Suppose that in this way we have proved that

$$a_0 = b_0, \quad a_1 = b_1, \ldots, \quad a_k = b_k.$$

Then (16.22) implies that

$$\begin{aligned} a_{k+1}(z_n - z_0)^{k+1} + a_{k+2}(z_n - z_0)^{k+2} + \cdots \\ = b_{k+1}(z_n - z_0)^{k+1} + b_{k+2}(z_n - z_0)^{k+2} + \cdots. \end{aligned} \tag{16.23}$$

Dividing (16.23) by $(z_n - z_0)^{k+1} \neq 0$, and passing to the limit as $n \to \infty$, we find that

$$a_{k+1} = b_{k+1}.$$

The proof now follows at once by induction.

80. The Relation between Power Series and Fourier Series

We now develop an analogy between power series and Fourier series. First we note that the polynomials $\{(z - z_0)^n\}$ are *orthogonal* in 0 on every circle γ_ρ with equation

$$z = \lambda(0) = z_0 + \rho e^{i\theta} \qquad (0 \leqslant 0 \leqslant 2\pi),$$

in the sense that[2]

$$\int_0^{2\pi} (z - z_0)^m \overline{(z - z_0)^n}\, d\theta = 0 \qquad \text{if} \quad m \neq n. \tag{16.24}$$

---

[2] Here we use $\int_0^{2\pi} \varphi(z)\, d0$ as an abbreviation for $\int_0^{2\pi} \varphi[\lambda(0)]\, d0$.

To prove (16.24), we merely note that the left-hand side equals

$$\int_0^{2\pi} (\rho e^{i\theta})^m (\rho e^{-i\theta})^n \, d\theta = \rho^{m+n} \int_0^{2\pi} e^{i(m-n)\theta} \, d\theta$$
$$= \rho^{m+n} \frac{e^{i(m-n)\theta}}{i(m-n)} \Big|_{\theta=0}^{\theta=2\pi} = 0 \qquad \text{if} \quad m - n \neq 0.$$

Next suppose ρ satisfies the condition $0 < \rho < R$, where R is the radius of convergence of the power series

$$f(z) = a_0 + a_1(z - z_0) + \cdots + a_m(z - z_0)^m + \cdots. \tag{16.25}$$

Since (16.25) is uniformly convergent on γ_ρ (see Theorem 16.3), so is the series

$$f(z)\overline{(z - z_0)^n} = a_0\overline{(z - z_0)^n} + a_1(z - z_0)\overline{(z - z_0)^n} + \cdots$$
$$+ a_m(z - z_0)^m\overline{(z - z_0)^n} + \cdots \qquad (n = 0, 1, 2, \ldots), \tag{16.26}$$

and hence term-by-term integration of (16.26) with respect to 0 along γ_ρ is permissible. The result is

$$\int_0^{2\pi} f(z)\overline{(z - z_0)^n} \, d\theta = a_n \int_0^{2\pi} (z - z_0)^n\overline{(z - z_0)^n} \, d\theta$$
$$= 2\pi\rho^{2n}a_n \qquad (n = 0, 1, 2, \ldots),$$

where we have used the *orthogonality relations* (16.24). Therefore

$$a_n = \frac{1}{2\pi\rho^{2n}} \int_0^{2\pi} f(z)\overline{(z - z_0)^n} \, d\theta \qquad (n = 0, 1, 2, \ldots), \tag{16.27}$$

and the original power series (16.25) can be written in the form

$$f(z) = \sum_{n=0}^{\infty} \frac{1}{2\pi\rho^{2n}} \int_0^{2\pi} f(z)\overline{(z - z_0)^n} \, d\theta \cdot (z - z_0)^n,$$

which explicitly represents $f(z)$ as an expansion with respect to the sequence of orthogonal functions $\{(z - z_0)^n\}$.

As another application of the orthogonality relations (16.24), we deduce that

$$\frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{k=0}^{n} a_k(z - z_0)^k \right|^2 d\theta = \frac{1}{2\pi} \int_0^{2\pi} \sum_{k=0}^{n} a_k(z - z_0)^k \sum_{l=0}^{n} \overline{a_l(z - z_0)^l} \, d\theta$$
$$= \frac{1}{2\pi} \int_0^{2\pi} \sum_{k,l=0}^{n} a_k\bar{a}_l(z - z_0)^k\overline{(z - z_0)^l} \, d\theta$$
$$= \frac{1}{2\pi} \sum_{k,l=0}^{n} a_k\bar{a}_l \int_0^{2\pi} (z - z_0)^k\overline{(z - z_0)^l} \, d\theta$$
$$= \sum_{k=0}^{n} |a_k|^2\rho^{2k},$$

bearing in mind that all integrals with $k \neq l$ vanish. Assuming that $\rho < R$, we now take the limit as $n \to \infty$ of the formula

$$\frac{1}{2\pi}\int_0^{2\pi}\left|\sum_{k=0}^{n} a_k(z - z_0)^k\right|^2 d0 = \sum_{k=0}^{n}|a_k|^2\rho^{2k} \tag{16.28}$$

just derived. The sequence of partial sums

$$\left\{\sum_{k=0}^{n} a_k(z - z_0)^k\right\}$$

is uniformly convergent on the circle $\gamma_\rho: |z - z_0| = \rho$, and each term of the sequence is continuous on γ_ρ. Moreover, according to Theorems 15.9, 15.10 and 15.11, the same is true of the sequence

$$\left\{\left|\sum_{k=0}^{n} a_k(z - z_0)^k\right|^2\right\},$$

and hence

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{2\pi}\int_0^{2\pi}\left|\sum_{k=0}^{n} a_k(z - z_0)^k\right|^2 d0 &= \frac{1}{2\pi}\int_0^{2\pi}\lim_{n\to\infty}\left|\sum_{k=0}^{n} a_k(z - z_0)^k\right|^2 d0 \\ &= \frac{1}{2\pi}\int_0^{2\pi}|f(z_0 + \rho e^{i0})|^2\, d0 \end{aligned} \tag{16.29}$$

(note that passing to the limit under the integral sign in (16.29) corresponds to term-by-term integration of an infinite series). It follows from (16.28) and (16.29) that the series

$$\sum_{n=0}^{\infty}|a_n|^2\rho^{2n}$$

converges, and that its sum is the integral in the right-hand side of (16.29). Thus we have proved the formula

$$\frac{1}{2\pi}\int_0^{2\pi}|f(z_0 + \rho e^{i0})|^2\, d0 = \sum_{n=0}^{\infty}|a_n|^2\rho^{2n}. \tag{16.30}$$

In the theory of Fourier series, this result is known as *Parseval's theorem*.[3]

Remark. According to (16.30), the integral

$$\frac{1}{2\pi}\int_0^{2\pi}|f(z_0 + \rho e^{i0})|^2\, d0,$$

representing the average of $|f(z)|^2$ on the circle γ_ρ, is a strictly increasing function of ρ [except for the trivial case where $f(z) \equiv \text{const}$]. Therefore the limit

$$\lim_{\rho\to R}\frac{1}{2\pi}\int_0^{2\pi}|f(z_0 + \rho e^{i0})|^2\, d0$$

always exists (but may be infinite).

---

[3] See e.g., G. P. Tolstov, *Fourier Series* (translated by R. A. Silverman), Prentice-Hall, Inc., Englewood Cliffs, N.J. (1962), p. 119.

It is instructive to cast formula (16.27) for the coefficients of a power series into a somewhat different form. Since $z = z_0 + \rho e^{i\theta}$, (16.27) becomes

$$a_n = \frac{1}{2\pi\rho^n}\int_0^{2\pi} f(z)e^{-in\theta}\,d\theta \qquad (n = 0, 1, 2, \ldots), \tag{16.31}$$

while, on the other hand, according to Cauchy's integral theorem,

$$\int_0^{2\pi} f(z)(z - z_0)^{n-1}\,dz = 0 \tag{16.32}$$

if $n \geqslant 1$. It follows from (16.32) that

$$\int_0^{2\pi} f(z)\rho^{n-1}\,e^{i(n-1)\theta}\,i\rho e^{i\theta}\,d\theta = 0,$$

or

$$0 = \frac{1}{2\pi\rho^n}\int_0^{2\pi} f(z)e^{in\theta}\,d\theta \qquad (n = 1, 2, \ldots). \tag{16.33}$$

Adding (16.31) to (16.33), and then subtracting (16.31) from (16.33), we obtain the formula

$$\begin{aligned} a_n &= \frac{1}{\pi\rho^n}\int_0^{2\pi} f(z)\cos n\theta\,d\theta \\ &= -\frac{i}{\pi\rho^n}\int_0^{2\pi} f(z)\sin n\theta\,d\theta \qquad (n = 1, 2, \ldots) \end{aligned} \tag{16.34}$$

which we supplement with the formula

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(z)\,d\theta, \tag{16.35}$$

obtained by setting $n = 0$ in (16.31).

Suppose now that we introduce the real and imaginary parts of a_n and $f(z)$, by writing

$$a_n = \alpha_n + i\beta_n, \qquad f(z) = u(\rho, \theta) + iv(\rho, \theta). \tag{16.36}$$

Then, substituting (16.36) into (16.34) and (16.35), and separating real and imaginary parts of the resulting formulas, we find that

$$\rho^n\alpha_n = \frac{1}{\pi}\int_0^{2\pi} u(\rho, \theta)\cos n\theta\,d\theta = \frac{1}{\pi}\int_0^{2\pi} v(\rho, \theta)\sin n\theta\,d\theta,$$

$$\rho^n\beta_n = \frac{1}{\pi}\int_0^{2\pi} v(\rho, \theta)\cos n\theta\,d\theta = -\frac{1}{\pi}\int_0^{2\pi} u(\rho, \theta)\sin n\theta\,d\theta, \qquad (n = 1, 2, \ldots),$$

and

$$\alpha_0 = \frac{1}{2\pi}\int_0^{2\pi} u(\rho, \theta)\,d\theta, \qquad \beta_0 = \frac{1}{2\pi}\int_0^{2\pi} v(\rho, \theta)\,d\theta.$$

Thus we see that the numbers α_0, $\rho^n\alpha_n$ and $-\rho^n\beta_n$ $(n = 1, 2, \ldots)$ are the Fourier coefficients of the function

$$u(\rho, \theta) = \mathrm{Re}\,f(z_0 + \rho e^{i\theta}),$$

while the numbers β_0, $\rho^n\beta_n$ and $\rho^n\alpha_n$ are the Fourier coefficients of the function

$$v(\rho, 0) = \operatorname{Im} f(z_0 + \rho e^{i\theta}).$$

In other words, the functions $u(\rho, 0)$ and $v(\rho, 0)$ have the formal Fourier expansions[4]

$$\begin{aligned} u(\rho, 0) &\sim \alpha_0 + \sum_{n=1}^{\infty} \rho^n(\alpha_n \cos n\theta - \beta_n \sin n\theta), \\ v(\rho, \theta) &\sim \beta_0 + \sum_{n=1}^{\infty} \rho^n(\beta_n \cos n\theta + \alpha_n \sin n\theta). \end{aligned} \tag{16.37}$$

In the theory of Fourier series, two such expansions are said to be *conjugates* of each other. The same series (16.37) can be obtained by separating real and imaginary parts of the convergent power series

$$\begin{aligned} f(z_0 + \rho e^{i\theta}) = u(\rho, 0) + iv(\rho, 0) &= \sum_{n=0}^{\infty} a_n(z - z_0)^n \\ &= \sum_{n=0}^{\infty} (\alpha_n + i\beta_n)\rho^n(\cos n\theta + i \sin n\theta), \end{aligned}$$

and, in fact, this alternative derivation shows that we can replace the sign $\sim$ by the sign $=$ in (16.37). In this way, the power series of a function $f(z)$ can be used to deduce the two conjugate Fourier series for the real and imaginary parts of $f(z)$.

Finally, we derive Parseval's theorem for the functions $u(\rho, 0)$ and $v(\rho, \theta)$ separately. Noting that

$$|f(z_0 + \rho e^{i\theta})|^2 = u^2(\rho, 0) + v^2(\rho, 0),$$

we can write (16.30) in the form

$$\frac{1}{2\pi}\int_0^{2\pi} u^2(\rho, 0)\, d\theta + \frac{1}{2\pi}\int_0^{2\pi} v^2(\rho, 0)\, d\theta = \sum_{n=0}^{\infty} |a_n|^2\rho^{2n}. \tag{16.38}$$

On the other hand, we can easily write a formula for the *difference* between the two integrals in the left-hand side of (16.38), by recalling that the average of the analytic function $f^2(z_0 + \rho e^{i\theta})$ on the circle $\gamma_\rho: |z - z_0| = \rho$ must coincide with the value of the function at the center of γ_ρ, i.e.,

$$\frac{1}{2\pi}\int_0^{2\pi} f^2(z_0 + \rho e^{i\theta})\, d\theta = f^2(z_0) \tag{16.39}$$

---

[4] As usual, the sign $\sim$ means that the Fourier series on the right is formed by using the Fourier coefficients of the function on the left, without knowing in advance that the formal Fourier series so obtained actually converges to the function on the left. See e.g., G. P. Tolstov, *op. cit.*, p. 14.

(see Theorem 14.2). Substituting the expressions

$$f^2(z_0 + \rho e^{i\theta}) = u^2(\rho, \theta) - v^2(\rho, \theta) + 2iu(\rho, \theta)v(\rho, \theta),$$
$$f^2(z_0) = a_0^2 = \alpha_0^2 - \beta_0^2 + 2i\alpha_0\beta_0$$

into (16.39), and taking the real part of the resulting equation, we find that

$$\frac{1}{2\pi}\int_0^{2\pi} u^2(\rho, \theta)\, d\theta - \frac{1}{2\pi}\int_0^{2\pi} v^2(\rho, \theta)\, d\theta = \alpha_0^2 - \beta_0^2. \tag{16.40}$$

Adding (16.40) to (16.38), and then subtracting (16.40) from (16.38), we obtain the formulas

$$\frac{1}{\pi}\int_0^{2\pi} u^2(\rho, \theta)\, d\theta = \alpha_0^2 - \beta_0^2 + \sum_{n=0}^{\infty} |a_n|^2\rho^{2n} = 2\alpha_0^2 + \sum_{n=1}^{\infty} (\alpha_n^2 + \beta_n^2)\rho^{2n},$$
$$\frac{1}{\pi}\int_0^{2\pi} v^2(\rho, \theta)\, d\theta = \beta_0^2 - \alpha_0^2 + \sum_{n=0}^{\infty} |a_n|^2\rho^{2n} = 2\beta_0^2 + \sum_{n=1}^{\infty} (\beta_n^2 + \alpha_n^2)\rho^{2n},$$

which expresses Parseval's theorem for the functions $u(\rho, \theta)$ and $v(\rho, \theta)$.

81. Expansion of an Analytic Function in a Power Series

In Sec. 79 we derived a formula for the coefficients of a power series in terms of the derivatives of the sum of the series at the center of its circle of convergence (of radius R):

$$a_n = \frac{f^{(n)}(z_0)}{n!} \qquad (n = 0, 1, 2, \ldots). \tag{16.41}$$

But, according to Theorem 14.7,

$$|f^{(n)}(z_0)| \leqslant n!\,\frac{M(\rho)}{\rho^n} \qquad (n = 0, 1, 2, \ldots) \tag{16.42}$$

for any $0 < \rho < R$, where

$$M(\rho) = \max_{z \in \gamma_\rho} |f(z)|$$

and γ_ρ is the circle $|z - z_0| = \rho$. Together, (16.41) and (16.42) imply the following *Cauchy inequalities* for the coefficients of a power series (often called the *Taylor coefficients*):

$$|a_n| \leqslant \frac{M(\rho)}{\rho^n} \qquad (n = 0, 1, 2, \ldots). \tag{16.43}$$

It is an immediate consequence of (16.43) that all the terms of the series

$$\sum_{n=0}^{\infty} |a_n|^2\rho^{2n} \tag{16.44}$$

are bounded, in fact, by the same constant $M^2(\rho)$. Actually, much more can

be said, since it follows at once from Parseval's theorem (16.30) and the obvious inequality

$$\frac{1}{2\pi}\int_0^{2\pi} |f(z_0 + \rho e^{i\theta})|^2\, d\theta \leqslant M^2(\rho)$$

that

$$\sum_{n=0}^{\infty} |a_n|^2 \rho^{2n} \leqslant M^2(\rho) \tag{16.45}$$

for all $\rho < R$, i.e., that the series (16.44) *itself* is bounded by $M^2(\rho)$. Now suppose the inequality (16.43) becomes an equality for some $n = N$ and for some ρ, $0 < \rho < R$, so that

$$|a_N| = \frac{M(\rho)}{\rho^N}$$

or

$$|a_N|^2 \rho^{2N} = M^2(\rho).$$

Then it follows from (16.45) that all the coefficients a_n with $n \neq N$ vanish, i.e., the power series

$$f(z) = a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots \tag{16.46}$$

reduces to the single term $a_N z^N$. Thus, except for the case where a power series reduces to a single one of its terms (with all the other terms vanishing), the sign $\leqslant$ can be replaced by the sign $<$ in (16.43). In particular, for $n = 0$ we find that

$$|a_0| = |f(z_0)| < M(\rho)$$

if $f(z) \not\equiv a_0$. The reader will recognize this result as a special case of Theorem 14.4, Corollary 2.

According to Theorem 14.5, Corollary 1, the derivatives of the sum of the power series (16.46) can be written in the form

$$f^{(n)}(z_0) = \frac{n!}{2\pi i}\int_{\gamma_\rho} \frac{f(z)}{(z - z_0)^{n+1}}\, dz \qquad (n = 0, 1, 2, \ldots) \tag{16.47}$$

if $\rho < R$. Therefore the coefficients of the power series can also be represented as integrals [cf. (16.41)]:

$$a_n = \frac{1}{2\pi i}\int_{\gamma_\rho} \frac{f(z)}{(z - z_0)^{n+1}}\, dz \qquad (n = 0, 1, 2, \ldots). \tag{16.48}$$

This last formula can easily be derived directly, without recourse to (16.47). In fact, multiplying the series (16.46) by the function

$$\frac{1}{2\pi i}\frac{1}{(z - z_0)^{n+1}},$$

we obtain a uniformly convergent series on γ_ρ (why?), which gives

$$\frac{1}{2\pi i}\int_{\gamma_\rho}\frac{f(z)}{(z-z_0)^{n+1}}\,dz=\sum_{m=0}^{\infty}a_m\frac{1}{2\pi i}\int_{\gamma_\rho}(z-z_0)^{m-n-1}\,dz, \tag{16.49}$$

when integrated term by term along γ_ρ. But

$$\begin{aligned}\int_{\gamma_\rho}(z-z_0)^{m-n-1}\,dz&=\int_0^{2\pi}\rho^{m-n-1}e^{i(m-n-1)\theta}i\rho e^{i\theta}\,d\theta\\&=i\rho^{m-n}\int_0^{2\pi}e^{i(m-n)\theta}\,d\theta,\end{aligned}$$

which vanishes if $m\neq n$, since then

$$i\rho^{m-n}\int_0^{2\pi}e^{i(m-n)\theta}\,d\theta=i\rho^{m-n}\frac{e^{i(m-n)\theta}}{i(m-n)}\Big|_{\theta=0}^{\theta=2\pi}=0,$$

and equals

$$i\int_0^{2\pi}d\theta=2\pi i$$

if $m=n$. Therefore all the terms in the right-hand side of (16.49) vanish, except the term corresponding to $m=n$, and (16.49) reduces to precisely (16.48).

We call attention once again to the fact that (16.48) was just deduced without starting from (16.47). This suggests that there is a way of deriving Cauchy's integral formula for a function $f(z)$ which is the sum of a power series without recourse to the argument used to prove Theorem 14.1. We now show that such a derivation is indeed possible. Substituting (16.48) into (16.46), we find that

$$f(z_1)=\sum_{n=0}^{\infty}\frac{1}{2\pi i}\int_{\gamma_\rho}\frac{f(z)}{(z-z_0)^{n+1}}\,dz\cdot(z_1-z_0)^n, \tag{16.50}$$

where z_1 is an arbitrary point of $I(\gamma_\rho)$. The series in the right-hand side of (16.50) can be obtained by integrating the series

$$\sum_{n=0}^{\infty}\frac{1}{2\pi i}\frac{f(z)}{(z-z_0)^{n+1}}(z_1-z_0)^n \tag{16.51}$$

term by term along γ_ρ, provided that (16.51) is uniformly convergent on γ_ρ. But on γ_ρ we have the estimate

$$\left|\frac{1}{2\pi i}f(z)\frac{(z_1-z_0)^n}{(z-z_0)^{n+1}}\right|\leqslant\frac{1}{2\pi}\frac{M(\rho)}{\rho}\left(\frac{|z_1-z_0|}{\rho}\right)^n, \tag{16.52}$$

and since the right-hand side of (16.52) is the general term of a convergent numerical series, i.e., of a geometric series with ratio

$$\frac{|z_1-z_0|}{\rho}<1,$$

the uniform convergence of (16.51) on γ_ρ follows by Weierstrass' M-test (Theorem 15.2). Therefore term-by-term integration of (16.51) along γ_ρ is permissible, and we find that

$$f(z_1) = \sum_{n=0}^{\infty} \frac{1}{2\pi i} \int_{\gamma_\rho} f(z) \frac{(z_1 - z_0)^n}{(z - z_0)^{n+1}}\, dz = \frac{1}{2\pi i} \int_{\gamma_\rho} f(z) \sum_{n=0}^{\infty} \frac{(z_1 - z_0)^n}{(z - z_0)^{n+1}}\, dz$$

$$= \frac{1}{2\pi i} \int_{\gamma_\rho} f(z) \frac{\dfrac{1}{z - z_0}}{1 - \dfrac{z_1 - z_0}{z - z_0}}\, dz = \frac{1}{2\pi i} \int_{\gamma_\rho} \frac{f(z)}{z - z_1}\, dz,$$

where we have summed the convergent geometric series

$$\sum_{n=0}^{\infty} \left(\frac{z_1 - z_0}{z - z_0} \right)^n$$

[recall that $z \in \gamma_\rho$, $z_1 \in I(\gamma_\rho)$]. Since z_1 is an arbitrary point of $I(\gamma_\rho)$, this is precisely Cauchy's integral formula for the function $f(z)$ [cf. Theorem 14.1].

To recapitulate, we can derive Cauchy's integral formula for the sum of the power series (16.46) if we first represent the coefficients a_n as integrals, then write the resulting sum of integrals as the integral of the uniformly convergent series (16.51), and finally sum this series behind the integral sign. By reversing this argument, Cauchy discovered that every function which can be represented as a Cauchy integral (i.e., every analytic function) can be expanded in a power series, thereby establishing the following remarkable result:

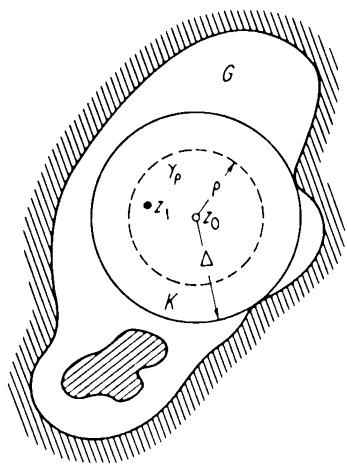

FIGURE 16.1

THEOREM 16.7 (*Cauchy's theorem on the expansion of an analytic function in a power series*). *Let $f(z)$ be an analytic function on a domain G, let z_0 be an arbitrary (finite) point of G, and let $\Delta = \Delta(z_0)$ be the distance between z_0 and the boundary of G.*[5] *Then there exists a power series*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \tag{16.53}$$

converging to $f(z)$ on the disk $K: |z - z_0| < \Delta$.

[5] If the boundary of G is empty or consists of the single point ∞, we write $\Delta = \infty$. Thus, if $f(z)$ is an entire function, the series (16.53) converges for all finite z.

Proof. Let z_1 be an arbitrary point of K, and let γ_ρ be any circle with center z_0 and radius ρ $(0 < \rho < \Delta)$ such that $z_1 \in I(\gamma_\rho)$ [see Figure 16.1]. Then, according to Cauchy's integral formula,

$$f(z_1) = \frac{1}{2\pi i}\int_{\gamma_\rho} \frac{f(z)}{z - z_1}\, dz.$$

To convert this relation into a power series of the form (16.53), we regard

$$\frac{1}{z - z_1}$$

as the sum of a geometric series with ratio

$$\frac{z_1 - z_0}{z - z_0},$$

where

$$\left|\frac{z_1 - z_0}{z - z_0}\right| = \frac{|z_1 - z_0|}{\rho} < 1.$$

To this end, we write

$$\begin{aligned}\frac{1}{z - z_1} &= \frac{1}{(z - z_0) - (z_1 - z_0)} = \frac{1}{z - z_0}\,\frac{1}{1 - \dfrac{z_1 - z_0}{z - z_0}} \\ &= \frac{1}{z - z_0} + \frac{z_1 - z_0}{(z - z_0)^2} + \cdots + \frac{(z_1 - z_0)^n}{(z - z_0)^{n+1}} + \cdots \\ &= \sum_{n=0}^{\infty} \frac{(z_1 - z_0)^n}{(z - z_0)^{n+1}}. \end{aligned} \tag{16.54}$$

We then multiply (16.54) by the function

$$\frac{1}{2\pi i} f(z),$$

obtaining

$$\frac{1}{2\pi i}\frac{f(z)}{z - z_1} = \sum_{n=0}^{\infty} \frac{1}{2\pi i}\frac{f(z)}{(z - z_0)^{n+1}}(z_1 - z_0)^n, \tag{16.55}$$

where the right-hand side is identical with (16.51), and is uniformly convergent on γ_ρ for the same reason. Therefore we can integrate (16.55) term by term along γ_ρ. The result is

$$f(z_1) = \frac{1}{2\pi i}\int_{\gamma_\rho} \frac{f(z)}{z - z_1}\, dz = \sum_{n=0}^{\infty} a_n (z_1 - z_0)^n,$$

where

$$a_n = \frac{1}{2\pi i}\int_{\gamma_\rho} \frac{f(z)}{(z - z_0)^{n+1}}\, dz = \frac{f^{(n)}(z_0)}{n!} \qquad (n = 0, 1, 2, \ldots).$$

Since the point $z_1 \in K$ is arbitrary, this (Taylor) series expansion is just the desired result (16.53).

Remark 1. Theorems 16.7 and 16.4, taken together, imply Theorem 1.2, cited without proof on p. 9. Theorem 1.2 is of fundamental importance, since it shows the equivalence between the definition of analyticity given in Chap. 1, in terms of the existence of *power series expansions*, and the definition given in Chap. 7, in terms of the existence of a *first derivative* (see p. 112). These two definitions have now been reconciled, i.e., we say that a function $f(z)$ is analytic at a point z_0 if $f(z)$ is differentiable in a neighborhood of z_0, or equivalently, if $f(z)$ has a convergent power series expansion in a neighborhood of z_0, and $f(z)$ is analytic on a domain G if it is analytic at every point of G. There is no such equivalence for functions of a real variable, or, more generally, for functions defined on a curve Γ (e.g., a Jordan curve in the extended plane). In fact, a function which is differentiable or even infinitely differentiable on Γ may not have a power series expansion in a neighborhood of any point of Γ, unless extra conditions are imposed on the function.[6]

Remark 2. According to Theorem 16.7, the Taylor series

$$f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n \tag{16.56}$$

representing the function $f(z)$ in a neighborhood of the point z_0 converges on the disk $|z - z_0| < \Delta$. Therefore the radius of convergence R of the series (16.56) can be no smaller than Δ, i.e.,

$$R \geqslant \Delta.$$

Writing this inequality in the form

$$\frac{1}{R} \leqslant \frac{1}{\Delta},$$

and substituting from the Cauchy-Hadamard formula (16.2), we find that

$$\overline{\lim_{n \to \infty}} \sqrt[n]{|a_n|} = \overline{\lim_{n \to \infty}} \sqrt[n]{\frac{|f^{(n)}(z_0)|}{n!}} \leqslant \frac{1}{\Delta}.$$

This will be recognized as the *Cauchy-Hadamard inequality*, already derived by another argument in Chap. 14 (see Theorem 14.8).

Example. Choosing $z_0 = 0$ in (16.56), and calculating the derivatives of all orders of the elementary functions e^z, $\cos z$, $\sin z$, $\cosh z$ and $\sinh z$, we immediately obtain the following power series expansions, valid for all finite z:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}, \quad \cos z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}, \quad \sin z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!},$$

$$\cosh z = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!}, \quad \sinh z = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!}.$$

---

[6] See Theorem 1.1 and Problem 1.9.

THEOREM 16.8 (*Liouville's theorem*). *If $f(z)$ is an entire function and if $|f(z)| \leqslant M$ for all finite z, then $f(z) \equiv$ const, i.e., every bounded entire function is a constant.*

Proof. Writing Cauchy's inequalities (16.43) for $n > 0$ only, we have

$$|a_n| \leqslant \frac{M(\rho)}{\rho^n} \leqslant \frac{M}{\rho^n} \qquad (n = 1, 2, \ldots) \tag{16.57}$$

for any $\rho > 0$. Taking the limit of (16.57) as $\rho \to \infty$, we find that $|a_n| \leqslant 0$, i.e., $a_n = 0$, for all $n = 1, 2, \ldots$. Therefore the power series (16.53) reduces to

$$f(z) \equiv a_0,$$

and the theorem is proved.

Remark. It follows from Liouville's theorem that any function which is analytic everywhere in the *extended plane* must be a constant. In fact, if $f(z)$ is analytic at infinity,[7] as well as for all finite z, then $f(z)$ is bounded for all z (why?).

Finally, we consider two examples of power series expansions of single-valued branches of multiple-valued functions:

Example 1. Let G be the domain with the half-line $x \leqslant 0$, $y = 0$ as its boundary (i.e., the whole plane cut along the negative real axis, including the origin), and let $\ln z$ denote the single-valued branch of the multiple-valued function $\operatorname{Ln} z$ which is defined on G and satisfies the condition $\ln 1 = 0$.[8] Then, since

$$\frac{d^n}{dz^n} \ln z = (-1)^{n-1} \frac{(n-1)!}{z^n} \qquad (n = 1, 2, \ldots),$$

it follows from (16.56), with $z_0 = 1$, that

$$\ln z = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{(z-1)^n}{n}. \tag{16.58}$$

The distance between the point $z_0 = 1$ and the boundary of G equals 1, and therefore the expansion (16.58) is valid for $|z - 1| < 1$. Replacing $z - 1$ by z, we obtain the following power series for $\ln(1 + z)$, called the *logarithmic series*, valid for $|z| < 1$:

$$\ln(1 + z) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{z^n}{n}.$$

---

[7] See Remark 1, p. 125.

[8] This branch can be represented in the form

$$\ln z = \ln |z| + i \arg z,$$

where $|\arg z| < \pi$.

Here $\ln(1 + z)$ is the single-valued branch of the function $\operatorname{Ln}(1 + z)$ which is defined on the domain G and satisfies the condition

$$\ln(1 + z)|_{z=0} = \ln 1 = 0.$$

Example 2. Consider the function

$$z^\alpha = \exp(\alpha \operatorname{Ln} z),$$

which is multiple-valued unless α is an integer (see Sec. 57). Let $\varphi(z)$ denote the branch of z^α which is defined on the same domain G as in Example 1, and satisfies the condition $\varphi(1) = 1$. This branch can be represented in the form

$$\varphi(z) = \exp(\alpha \ln z). \tag{16.59}$$

Therefore

$$\varphi'(z) = \alpha \exp(\alpha \ln z)\frac{1}{z} = \alpha \exp(\alpha \ln z) \exp(-\ln z) = \alpha \exp[(\alpha - 1) \ln z],$$

and similarly,

$$\varphi^{(n)}(z) = \alpha(\alpha - 1) \cdots (\alpha - n + 1) \exp[(\alpha - n) \ln z] \qquad (n = 1, 2, \ldots).$$

It follows from (16.56), with $z_0 = 1$, that

$$z^\alpha = \sum_{n=0}^{\infty} \frac{\alpha(\alpha - 1) \cdots (\alpha - n + 1)}{n!}(z - 1)^n,$$

where we replace $\varphi(z)$ by z^α, with the understanding that z^α is the single-valued function (16.59) defined on the domain G.[9] Just as in Example 1, this expansion is valid for $|z - 1| < 1$. Replacing $z - 1$ by z, we obtain the following power series for $(1 + z)^\alpha$, called the *binomial series*, valid for $|z| < 1$:

$$(1 + z)^\alpha = \sum_{n=0}^{\infty} \frac{\alpha(\alpha - 1) \cdots (\alpha - n + 1)}{n!} z^n. \tag{16.60}$$

Formula (16.60) is the appropriate generalization of the familiar *binomial theorem* to the case where the exponent α is an arbitrary complex number.

PROBLEMS

16.1. Prove that if a power series is absolutely convergent at a single point of its circle of convergence, then it is absolutely convergent at every point of its circle of convergence.

[9] The coefficient of $(z - 1)^n$ is defined as 1 if $n = 0$.

16.2. Find the radius of convergence of each of the following power series:

a) $\sum_{n=1}^{\infty} n^k z^n$ $(k = 0, 1, 2, \ldots)$; b) $\sum_{n=1}^{\infty} n^n z^n$; c) $\sum_{n=1}^{\infty} \frac{n^k}{n!} z^n$ $(k = 0, 1, 2, \ldots)$;

d) $\sum_{n=1}^{\infty} \frac{n!}{n^n} z^n$; e) $\sum_{n=0}^{\infty} \frac{2^n}{n!} z^n$; f) $\sum_{n=1}^{\infty} n! z^n$; g) $\sum_{n=1}^{\infty} \frac{z^n}{n}$; h) $\sum_{n=1}^{\infty} \frac{n}{2^n} z^n$; i) $\sum_{n=0}^{\infty} z^{n!}$;

j) $\sum_{n=0}^{\infty} 2^n z^{n!}$; k) $\sum_{n=0}^{\infty} z^{2^n}$; l) $\sum_{n=0}^{\infty} [3 + (-1)^n]^n z^n$; m) $\sum_{n=0}^{\infty} (\cos in) z^n$;

n) $\sum_{n=0}^{\infty} (n + a^n) z^n$.

Ans. b) 0; d) e; f) 0; h) 2; j) 1;
l) $\frac{1}{4}$; n) 1 if $|a| \leq 1$, $1/|a|$ if $|a| > 1$.

16.3. Given that the radius of convergence of the power series

$$\sum_{n=0}^{\infty} c_n z^n$$

is R $(0 < R < \infty)$, what is the radius of convergence of each of the following series:

a) $\sum_{n=0}^{\infty} n^k c_n z^n$ $(k = 0, 1, 2, \ldots)$; b) $\sum_{n=0}^{\infty} (2^n - 1) c_n z^n$; c) $\sum_{n=0}^{\infty} \frac{c_n}{n!} z^n$;

d) $\sum_{n=1}^{\infty} n^n c_n z^n$; e) $\sum_{n=0}^{\infty} c_n^k z^n$ $(k = 0, 1, 2, \ldots)$; f) $\sum_{n=0}^{\infty} (1 + z_0^n) c_n z^n$?

Ans. a) R; c) ∞; e) R^k.

16.4. Given that the radii of convergence of the power series

$$\sum_{n=0}^{\infty} a_n z^n \quad \text{and} \quad \sum_{n=0}^{\infty} b_n z^n$$

equal R_1 and R_2, respectively, what can be said about the radius of convergence R of each of the following series:

a) $\sum_{n=0}^{\infty} (a_n + b_n) z^n$; b) $\sum_{n=0}^{\infty} (a_n - b_n) z^n$; c) $\sum_{n=0}^{\infty} a_n b_n z^n$; $\sum_{n=0}^{\infty} \frac{a_n}{b_n} z^n$ $(b_n \neq 0)$?

Ans. a) $R \geq \min(R_1, R_2)$; c) $R \geq R_1 R_2$.

16.5. Give an example of two power series

$$\sum_{n=0}^{\infty} a_n z^n, \qquad \sum_{n=0}^{\infty} b_n z^n$$

with the same finite radius of convergence, such that the radius of convergence of the series

$$\sum_{n=0}^{\infty} (a_n + b_n) z^n$$

is infinite.

16.6. Prove that series expansions of the form

$$a_0 \varphi_0(x, y) + a_1 \varphi_1(x, y) + \cdots + a_n \varphi_n(x, y) + \cdots,$$

where

$$\varphi_0(x, y) = 1, \quad \varphi_1(x, y) = x, \quad \varphi_2(x, y) = x^2 - y^2, \quad \varphi_3(x, y) = x^3 - 3xy^2, \ldots,$$

have the uniqueness property (on suitable sets). What can be said about the sum $f(x, y)$ of a uniformly convergent expansion of this type?

Hint. $\varphi_n(x, y) = \operatorname{Re} z^n = \operatorname{Re} (x + iy)^n$.

Ans. $f(x, y)$ is harmonic (see Problem 14.15).

16.7. Verify the following Fourier expansions:

$$e^{\cos\theta} \cos(\sin\theta) = 1 + \cos\theta + \frac{\cos 2\theta}{2!} + \cdots + \frac{\cos n\theta}{n!} + \cdots,$$

$$e^{\cos\theta} \sin(\sin\theta) = \sin\theta + \frac{\sin 2\theta}{2!} + \cdots + \frac{\sin n\theta}{n!} + \cdots.$$

16.8. Show that

a) $\dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} e^{2\cos\theta}\, d\theta = \sum_{n=0}^{\infty} \frac{1}{(n!)^2}$;

b) $\dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} e^{2\cos\theta} \cos(2\sin\theta)\, d\theta = 2\pi$.

16.9. Prove the formula

$$\frac{1}{1 - \rho^2} = \frac{1}{2\pi}\int_0^{2\pi} \frac{d\theta}{1 - 2\rho\cos\theta + \rho^2} \qquad (0 \leqslant \rho < 1).$$

Hint. Use formula (16.30).

16.10. Verify the following Fourier expansions, where $\rho > 1$:

a) $\dfrac{\rho(\rho - \cos\theta)}{1 - 2\rho\cos\theta + \rho^2} = 1 + \dfrac{\cos\theta}{\rho} + \dfrac{\cos 2\theta}{\rho^2} + \cdots + \dfrac{\cos n\theta}{\rho^n} + \cdots$;

b) $\dfrac{\rho\sin\theta}{1 - 2\rho\cos\theta + \rho^2} = \dfrac{\sin\theta}{\rho} + \dfrac{\sin 2\theta}{\rho^2} + \cdots + \dfrac{\sin n\theta}{\rho^n} + \cdots$.

16.11. Give an alternative proof of Liouville's theorem (Theorem 16.8), based on the use of Cauchy's integral formula.

Hint. Consider the integral

$$\int_{|z| = R} \frac{f(z)}{(z - a)(z - b)}\, dz,$$

where $a \neq b$, $|a| < R$, $|b| < R$, and take the limit as $R \to \infty$.

16.12. Prove the following generalization of Liouville's theorem: If $f(z)$ is an entire function, and if the function

$$M(\rho) = \max_{|z| = \rho} |f(z)|$$

satisfies the inequality

$$M(\rho) \leqslant M\rho^k,$$

where M is a positive constant and k is a fixed positive integer, then $f(z)$ is a polynomial of degree no higher than k.

16.13. Sum the following power series for $|z| < 1$:

a) $\sum_{n=1}^{\infty} \frac{z^n}{n}$; b) $\sum_{n=0}^{\infty} \frac{z^{2n+1}}{2n+1}$.

Ans. a) $-\ln(1-z)$.

16.14. Expand each of the following functions in a Taylor series at the point $z = 0$:

a) $\sin^2 z$; b) $\cosh^2 z$; c) $\sqrt{z+i}$, where $\sqrt{i} = \frac{1+i}{\sqrt{2}}$;

d) $\frac{1}{az+b}$, where $b \neq 0$; e) $\frac{z}{z^2-4z+13}$; f) $\frac{z^2}{(z+1)^2}$;

g) arc tan z, where arc tan $0 = 0$; h) $\int_0^z e^{\zeta^2}\,d\zeta$; i) $\int_0^z \frac{\sin\zeta}{\zeta}\,d\zeta$.

In each case, find the radius of convergence of the Taylor series.

16.15. Expand each of the following functions in a Taylor series at the point $z = 1$:

a) $\frac{z}{z+2}$; b) $\frac{z^2}{(z+1)^2}$; c) $\sqrt[3]{z}$, where $\sqrt[3]{1} = \frac{-1+i\sqrt{3}}{2}$;

d) $\frac{z}{z^2-2z+5}$.

In each case, find the radius of convergence of the Taylor series.

17

POWER SERIES: RAMIFICATIONS

82. The Interior Uniqueness Theorem. A-Points of an Analytic Function

Suppose two power series in the variable $z - z_0$ have the same sum on a set E which has z_0 as a limit point. Then, according to Theorem 16.6, the two series have the same coefficients and the same radius of convergence R, and hence their sums coincide everywhere on the disk $K: |z - z_0| < R$. In other words, the sum of a power series is determined on the whole disk K once its values are known on any set of points with the center of the disk as a limit point. Since a power series represents an analytic function on K (see Theorem 16.4), this prompts us to ask the following question: If $f(z)$ is an analytic function on a domain G, is $f(z)$ completely determined by its values on an arbitrary set $E \subset G$ which has at least one limit point in G? Equivalently, do the values of two analytic functions $f(z)$ and $g(z)$ defined on the same domain G coincide at all points of G if they coincide on a set E with a limit point in G? As we now show, the answer to these questions is in the affirmative:

THEOREM 17.1 (*Interior uniqueness theorem*). *Suppose two single-valued analytic functions $f(z)$ and $g(z)$, defined on a domain G, coincide on a set $E \subset G$ with a limit point $z_0 \in G$. Then $f(z)$ and $g(z)$ coincide on the whole domain G.*

Proof. Let

$$a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots$$

and

$$b_0 + b_1(z - z_0) + \cdots + b_n(z - z_0)^n + \cdots$$

be the power series representing $f(z)$ and $g(z)$, respectively, in a neighborhood $\mathcal{N}(z_0)$: $|z - z_0| < \rho$, $\mathcal{N}(z_0) \subset G$. Since the sums of these two series coincide on E, it follows from Theorem 16.6 that $f(z)$ and $g(z)$ coincide on $\mathcal{N}(z_0)$. Let E_0 denote the subset of G consisting of all points with the same property as z_0, i.e., all points which have neighborhoods in which $f(z)$ and $g(z)$ coincide. Clearly, E_0 is a nonempty open set. Moreover, if $\zeta \in G$ is a limit point of E_0, we can apply the same argument to ζ and E_0 as just applied to z_0 and E, and hence ζ must belong to E_0. Therefore, according to Theorem 4.16, $E_0 = G$, and the theorem is proved.

Remark 1. It follows from Theorem 17.1 that if $f(z)$ and $g(z)$ coincide on some subdomain $D \subset G$ (in particular, in a neighborhood of a point $z_0 \in G$), or on an arc of some continuous curve contained in G, then $f(z)$ and $g(z)$ coincide on G itself. In particular, if $f(z)$ is analytic on a domain G and if $f(z)$ is not identically equal to a constant on G, then $f(z)$ cannot have a constant value in any neighborhood of a point $z_0 \in G$ or on any arc of a curve $\gamma \subset G$. Moreover, if $f(z) \not\equiv \text{const}$ on G, and if z_0 is an arbitrary point of G, then at least one of the derivatives $f'(z_0), \ldots, f^{(n)}(z_0), \ldots$ must be nonzero, since otherwise the Taylor series expansion of $f(z)$ would have the form

$$f(z) = a_0 + 0 \cdot (z - z_0) + \cdots + 0 \cdot (z - z_0)^n + \cdots$$

in a neighborhood of z_0, which implies $f(z) \equiv f(z_0)$ for all $z \in G$.

Remark 2. Besides the interior uniqueness theorem just proved, there are various uniqueness theorems involving *boundary values* of analytic functions. Before stating one such theorem, we introduce a few definitions: Let E be a bounded set of points on the real line. Then by the *outer measure* of E we mean the quantity

$$\mu^*(E) = \sup \sum_{n=1}^{\infty} (b_n - a_n),$$

where the least upper bound is over all countable sets of intervals $(a_1, b_1), \ldots, (a_n, b_n), \ldots$, with $a_n < b_n$, such that

$$E \subset (a_1, b_1) \cup \cdots \cup (a_n, b_n) \cup \cdots$$

(cf. Definition 1, p. 310, and Problem 17.4). If $I = (a, b)$ contains E, then by the *inner measure* of E we mean the quantity

$$\mu_*(E) = (b - a) - \mu^*(I - E).$$

If $\mu_*(E) = \mu^*(E)$, and if we denote this common value by $\mu(E)$, then E is said to be *measurable*, with *measure* $\mu(E)$. The measure of sets of points on a rectifiable curve $\Gamma: z = \lambda(t)$ is defined in the same way, except that now "open arcs" (i.e., arcs which are the images under $\lambda(t)$ of open intervals) and

their lengths play the role of the open intervals (a_k, b_k) and their lengths $b_k - a_k$.

We now cite without proof the following "boundary uniqueness theorem," due to Luzin and Privalov: *Let G be the interior of a closed rectifiable Jordan curve Γ, and let $f(z)$ and $g(z)$ be two single-valued analytic functions defined on G. Suppose $E \subset \Gamma$ is a set of positive measure such that the difference $f(z) - g(z)$ has a zero limit at every point $z_0 \in E$ as a variable point $z \in G$ approaches z_0 along a nontangential path (cf. Definition 2, p. 311).*[1] *Then $f(z)$ and $g(z)$ coincide on the whole domain G.*

For a proof of this theorem, requiring a knowledge of real variable theory, we refer the reader to the book by Privalov cited on p. 316. There he will find other theorems of this character, in particular, a uniqueness theorem involving *radial boundary values*, i.e., limits as $z \in G$ approaches $z_0 \in E$ along the *normal* to Γ at z_0 (along the radius, in the case where Γ is a circle).

By an *A-point* of a function $f(z)$, we mean a root of the equation $f(z) = A$, where A is an arbitrary finite complex number. It follows from Theorem 17.1 that *if $f(z) \not\equiv A$ is analytic on a domain G, the set of A-points of $f(z)$ cannot have a limit point in G.* Moreover, any compact subset $F \subset G$ can only contain a finite number of A-points (for fixed A), since if F contained infinitely many A-points, these points would have a limit point belonging to F and hence to G. If $z_0 \in G$ is any A-point of $f(z)$, then $f(z_0) = A$, and hence, making a power series expansion of $f(z)$ in a neighborhood of z_0, we have

$$f(z) = A + f'(z_0)(z - z_0) + \frac{f''(z_0)}{2!}(z - z_0)^2 + \cdots$$

or

$$f(z) - A = f'(z_0)(z - z_0) + \frac{f''(z_0)}{2!}(z - z_0)^2 + \cdots. \tag{17.1}$$

If $f(z) \not\equiv$ const, then, according to Remark 1 above, at least one of the coefficients $f'(z_0), f''(z_0), \ldots$ is nonzero. Let $(z - z_0)^k$ be the smallest power of $z - z_0$ with a nonzero coefficient. Then (17.1) becomes

$$f(z) - A = (z - z_0)^k \left[\frac{f^{(k)}(z_0)}{k!} + \frac{f^{(k+1)}(z_0)}{(k+1)!}(z - z_0) + \cdots\right], \tag{17.2}$$

where $f^{(k)}(z_0) \neq 0$. The integer $k \geqslant 1$ is called the *order* of the A-point z_0, and z_0 is called a *simple* A-point if $k = 1$ and a *multiple* A-point if $k > 1$. It follows from (17.2) and this definition that

$$f(z_0) = A, \qquad f'(z_0) \neq 0$$

if z_0 is a simple A-point, whereas

$$f(z_0) = A, \quad f'(z_0) = 0, \ldots, \quad f^{(k-1)}(z_0) = 0, \quad f^{(k)}(z_0) \neq 0 \tag{17.3}$$

---

[1] This presupposes that Γ has a tangent at every point $z_0 \in E$, but this is no restriction, in view of Lemma 2′, p. 310.

if z_0 is an A-point of order k. From now on, in determining the number of A-points in a compact subset $F \subset G$, we shall count each A-point a number of times equal to its order. Then the number of A-points in F will still be finite, since F contains only a finite number of multiple A-points and each is of finite order.

THEOREM 17.2. *If $f(z)$ is analytic at z_0, then $f(z)$ has an A-point of order k at z_0 if and only if the function*

$$\varphi(z) = \frac{f(z) - A}{(z - z_0)^k} \tag{17.4}$$

is analytic at z_0 and

$$\varphi(z_0) = \lim_{z \to z_0} \varphi(z) \neq 0. \tag{17.5}$$

Proof. If $f(z)$ has an A-point of order k at z_0, then, according to (17.2) and (17.4), $\varphi(z)$ has the power series expansion

$$\varphi(z) = \frac{f^{(k)}(z_0)}{k!} + \frac{f^{(k+1)}(z_0)}{(k+1)!}(z - z_0) + \cdots$$

in a neighborhood of z_0, where

$$\varphi(z_0) = \frac{f^{(k)}(z_0)}{k!} \neq 0$$

[cf. (17.3)]. Conversely, if $\varphi(z)$ is analytic at z_0 and the condition (17.5) holds, then

$$\varphi(z) = a_k + a_{k+1}(z - z_0) + \cdots$$

in a neighborhood of z_0, where $a_k = \varphi(z_0) \neq 0$. Substituting for $\varphi(z)$ from (17.4) and solving for $f(z)$, we find that

$$f(z) = A + a_k(z - z_0)^k + a_{k+1}(z - z_0)^{k+1} + \cdots.$$

Obviously the conditions (17.3) hold, i.e., $f(z)$ has an A-point of order k at z_0.

Remark. Everything just said about A-points applies equally to the special case where $A = 0$. For simplicity, *zero-points* are usually called *zeros*. The reader can easily verify that our present definition of the order of a zero is consistent with the definition already given in the case where $f(z)$ is a polynomial (see p. 129) or a rational function (see p. 161). Since the A-points of the function $f(z)$ are the zeros of the function $f(z) - A$, any general property which we prove for the zeros of an analytic function (e.g., Theorem 17.3 below) is also valid for the A-points of an analytic function.

DEFINITION. *If the function $f(z)$ is analytic at ∞ (see p. 125), and if $f(\infty) = A$, we say that $f(z)$ has an A-point of order k at ∞ if $f^*(\zeta) = f(1/\zeta)$ has an A-point of order k at $\zeta = 0$.*

THEOREM 17.3. *Let $f(z)$ be a single-valued analytic function defined on a domain G. Then the set of all zeros of $f(z)$ in G is countable.*

Proof. Obviously, there is no loss of generality in assuming that G does not contain the point at infinity. Since the union of countably many countable sets is countable (see Problem 3.2), and since any compact set contains only finitely many zeros of $f(z)$, the theorem will be proved if we show that G can always be represented as a union of countably many compact sets. If G is the whole plane, G is (trivially) the union of the closed disks

$$|z| \leqslant 1, \quad |z| \leqslant 2, \ldots, \quad |z| \leqslant n, \ldots$$

On the other hand, if Γ, the boundary of G, contains at least one finite point, let $\rho(z, \Gamma)$ denote the distance between an arbitrary point z and Γ (see p. 52). Then let F_n denote the set of all points in G such that

$$|z| \leqslant n, \qquad \rho(z, \Gamma) \geqslant \frac{1}{n}.$$

Clearly, all the F_n starting from some value of n are nonempty, and every point of G belongs to all the F_n starting from some value of n. It follows that

$$G = F_1 \cup \cdots \cup F_n \cup \cdots \tag{17.6}$$

Since the F_n are obviously bounded, it only remains to show that the F_n are closed. If F_n is empty, F_n is closed by definition (see p. 47). Every nonempty F_n is also closed, since if ζ is a limit point of F_n, then ζ is also a limit point of G, and moreover

$$\rho(\zeta, \Gamma) = \lim_{\substack{z \to \zeta \\ z \in F_n}} \rho(z, \Gamma) \geqslant \frac{1}{n},$$

which implies $\zeta \in G$, $\zeta \in F_n$. Here we use the easily verified fact that $\rho(z, \Gamma)$ is a continuous function of z (cf. p. 54 and Problem 3.28). This completes the proof.

Remark. Because of (17.6) and the fact that $F_n \subset F_{n+1}$ for all $n = 1, 2, \ldots$, we call $\{F_n\}$ an *increasing sequence of compact sets approximating* G.

Examination of the power series

$$f(z) = a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots$$

tells us immediately whether or not z_0 is a zero of the function $f(z)$, as well as the order of z_0, provided z_0 is indeed a zero, but it does not tell us whether or not $f(z)$ has zeros other than z_0, where such zeros are located, and what their orders are. For example, the function $\sin z$ can be represented as a power series

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots, \tag{17.7}$$

which converges for all finite z, and it is obvious from (17.7) that $\sin z$ has

a simple zero at $z = 0$. However, (17.7) does not reveal that $\sin z$ has simple zeros at all the points $\pm\pi$, $\pm 2\pi, \ldots$, and that it has no zeros other than those just mentioned. Actually, infinite products rather than power series constitute the appropriate tool for representing analytic functions in a way which reveals all their zeros in a given domain.

LEMMA. *Given a sequence $\{f_n(z)\}$ of analytic functions defined on a domain G, let E_n be the set of all zeros of $f_n(z)$ and let E be the union of all the sets E_n $(n = 1, 2, \ldots)$. Suppose G contains no limit points of E, and suppose every point of E is a zero of only finitely many functions $f_n(z)$. Then, given an arbitrary compact subset $F \subset G$, F contains only finitely many points of E, and there exists an integer $N(F) \geq 0$ such that $f_n(z)$ has no zeros in F if $n > N(F)$.*[2]

Proof. If F contains infinitely many points of E, then, since F is compact, F contains a limit point of E, i.e., G contains a limit point of E, contrary to hypothesis. Therefore F contains only finitely many points of E. Since each of these points can only be a zero of finitely many $f_n(z)$, the existence of an integer $N(F) \geq 0$ with the desired property is obvious.

THEOREM 17.4. *Given a domain G and an arbitrary compact subset $F \subset G$, let $\{f_n(z)\}$, E and $N(F)$ be the same as in the lemma, and suppose the series*

$$\sum_{n=N(F)+1}^{\infty} \ln f_n(z) \tag{17.8}$$

is uniformly convergent on F. Then the infinite product

$$\prod_{n=1}^{\infty} f_n(z)$$

converges uniformly to an analytic function $f(z)$ on every compact subset of G. Moreover, $f(z)$ has a zero at an arbitrary point $z_0 \in G$ if and only if at least one of the functions $f_n(z)$ has a zero at z_0.

Proof. Here, of course, we must use the extended definition of infinite products given at the end of Sec. 77. As in the proof of Theorem 15.15, uniform convergence of the series (17.8) on F implies uniform convergence of the infinite product

$$\prod_{n=N(F)+1}^{\infty} f_n(z) \tag{17.9}$$

on F. The function

$$\varphi(z) = \begin{cases} \displaystyle\prod_{n=1}^{N(F)} f_n(z) & \text{if } N(F) > 0, \\ 1 & \text{if } N(F) = 0 \end{cases}$$

[2] $N(F) = 0$ if *no* $f_n(z)$ has zeros in F.

is continuous and hence bounded on F. It follows from Theorem 15.9 that the sequence of partial products

$$\{p_n(z)\} = \{f_1(z) \cdots f_n(z)\}$$

is uniformly convergent on F, since

$$p_n(z) = \varphi(z)q_n(z) \equiv \varphi(z)f_{N(F)+1} \cdots f_n(z)$$

if $n > N(F)$. The rest of the proof is now a consequence of Theorem 15.8 and the fact that (17.9) is not allowed to have a zero limit.

Remark 1. Only finitely many of the functions $f_n(z)$ can have a zero at a given point $z_0 \in E \subset G$. Let these functions be denoted by

$$f_{n_1}(z), \ldots, f_{n_m}(z),$$

and let $\alpha_1, \ldots, \alpha_m$ be the corresponding orders of the zero at z_0. Then

$$f(z) = \prod_{n=1}^{\infty} f_n(z)$$

has a zero of order

$$\alpha = \alpha_1 + \cdots + \alpha_m$$

at z_0. To see this, we write

$$f(z) = \prod_{n=1}^{\infty} f_n(z) = \prod_{n=1}^{\nu} f_n(z) \cdot \prod_{n=\nu+1}^{\infty} f_n(z),$$

where

$$\nu = \max(n_1, \ldots, n_m),$$

and note that the first factor on the right is an analytic function with a zero of order α at z_0 (see Problem 17.5), whereas the second factor cannot vanish at z_0.

Remark 2. Suppose every zero of $f_n(z)$ is known, e.g., suppose every $f_n(z)$ can be written in the form

$$f_n(z) = (z - z_n)g_n(z),$$

where every $g_n(z)$ is an analytic function with no zeros in G. Then we have the representation

$$f(z) = \prod_{n=1}^{\infty} f_n(z) = \prod_{n=1}^{\infty} (z - z_n)g_n(z), \tag{17.10}$$

from which we can immediately read off all the zeros of $f(z)$ and their orders. In fact, the order of z_n is just the number of times the linear factor $z - z_n$ appears in the right-hand side of (17.10).

83. The Maximum Modulus Principle and Some of its Consequences. Lemniscates

In Sec. 71 we derived a preliminary version of the maximum modulus principle. We are now in a position to prove this principle in its final form:

THEOREM 17.5 (*Maximum modulus principle*). *If $f(z)$ is an analytic function on a domain G, and if G contains the circle $\gamma_\rho: |z - z_0| = \rho$ and its interior, then*

$$|f(z_0)| < M(\rho) = \max_{z \in \gamma_\rho} |f(z)|,$$

provided that $f(z) \not\equiv$ const on G. Thus the modulus of $f(z)$ cannot have a maximum[3] *at any point of G.*

Proof. The only difference between the first assertion of Theorem 17.5 and the first assertion of Theorem 14.4, Corollary 2 is that we now assume $f(z) \not\equiv$ const on G, instead of $f(z) \not\equiv$ const on $I(\gamma_\rho)$. Therefore, the theorem will be proved if we can show that $f(z) \not\equiv$ const on G implies $f(z) \not\equiv$ const in every neighborhood contained in G. But this follows at once from the interior uniqueness theorem (see Remark 1, p. 370).

COROLLARY (*Minimum modulus principle*). *The modulus of an analytic function $f(z) \not\equiv$ const, defined on a domain G, cannot have a minimum at any point of G which is not a zero of $f(z)$.*

Proof. If $f(z_0) \neq 0$, then, since $f(z)$ is continuous, the inequality $f(z) \neq 0$ holds in some neighborhood $\mathcal{N}(z_0) \subset G$. Therefore the function

$$\varphi(z) = \frac{1}{f(z)}$$

is analytic and not identically a constant on $\mathcal{N}(z_0)$. It follows from Theorem 17.5 that $|\varphi(z)|$ cannot have a maximum at z_0, i.e., $|f(z)|$ cannot have a minimum at z_0. Obviously, $|f(z)|$ has a minimum at any point z_0 which is a zero of $f(z)$.

THEOREM 17.6. *Given a bounded domain G, let $f(z) \not\equiv$ const be a function which is analytic on G and continuous on $\overline{G}$, and suppose $|f(z)|$ has the same value at all points of the boundary of G. Then $f(z)$ has at least one zero at a point of G.*

Proof. Suppose $f(z)$ has no zeros in G. Then, according to Theorem 17.5 and its corollary, $|f(z)|$ can have neither a maximum nor a minimum

[3] More precisely, a relative maximum in the wide sense (see footnote 3, p. 296).

in G. Let Γ denote the boundary of G. Since $f(z)$ is continuous on the compact set $\bar{G} = G \cup \Gamma$, $|f(z)|$ must achieve both its maximum and its minimum on $\bar{G}$ (see Theorem 3.16), and hence on Γ. But by hypothesis $|f(z)| \equiv \text{const}$ on Γ. Therefore

$$\min_{z \in \bar{G}} |f(z)| = \max_{z \in \bar{G}} |f(z)|,$$

i.e., $|f(z)| \equiv \text{const}$ on $\bar{G}$, and hence $f(z) \equiv \text{const}$, by Theorem 17.5. This contradiction completes the proof.

THEOREM 17.7 (*Fundamental theorem of algebra*). *Every polynomial*

$$P_n(z) = a_0 + a_1 z + \cdots + a_n z^n \qquad (a_n \neq 0, \quad n > 0)$$

has a zero in the complex plane.

Proof. If $a_0 = 0$, then $P_n(0) = 0$, and the theorem is trivial. Therefore, suppose $a_0 \neq 0$, so that $P_n(0) \neq 0$. Consider the set E of all points z such that $|P_n(z)| < 2|a_0|$. E is nonempty, since E contains the point $z = 0$, and E is bounded, since $|P_n(z)|$ can be made arbitrarily large (by suitably increasing $|z|$), in particular, greater than $2|a_0|$. Moreover, E is open, since if $|P_n(z_0)| < 2|a_0|$ for $z_0 \in E$, then, by the continuity of $|P_n(z)|$, we have $|P_n(z)| < 2|a_0|$ for all z in some neighborhood $\mathcal{N}(z_0)$, i.e., $\mathcal{N}(z_0) \subset E$. Finally, $|P_n(z)| = 2|a_0|$ on the boundary of E; to see this, we note that if ζ is a boundary point of E, then $|P_n(\zeta)| \geqslant 2|a_0|$, since $\zeta \notin E$, while on the other hand, $|P_n(\zeta)| \leqslant 2|a_0|$, since every neighborhood of ζ contains points for which $|P_n(z)| < 2|a_0|$. Being an open set, E is the union of countably many disjoint domains (see Theorem 4.7). Applying Theorem 17.6 to each of these domains, we find that each contains at least one zero of the polynomial $P_n(z)$. This proves the theorem.

Remark. By the usual argument (see p. 129), it follows from Theorem 17.7 that $P_n(z)$ has precisely n zeros, where each multiple zero is counted a number of times equal to its order. In particular, E cannot be the union of more than n disjoint domains.

By precisely the same reasoning as given in the proof of Theorem 17.7, we see that the set E_ρ of all points z such that $|P_n(z)| < \rho \, (0 < \rho < \infty)$ is nonempty, bounded and open.[4] Moreover, E_ρ contains all the zeros of $P_n(z)$, $|P_n(z)| = \rho$ on the boundary of E_ρ, and E_ρ consists of at most n disjoint domains, each of which contains one or more zeros of $P_n(z)$. Let G_ρ be any of these domains, and let γ be an arbitrary closed Jordan curve contained in G_ρ. Since $|P_n(z)| < \rho$ for all $z \in \gamma$, it follows from the maximum modulus principle that $|P_n(z)| < \rho$ for all $z \in I(\gamma)$. Therefore $I(\gamma) \subset E_\rho$ and $I(\gamma) \subset G_\rho$,

[4] The choice $\rho = 2|a_0|$ made in the proof of Theorem 17.7 guarantees that E is nonempty, before it is known that $P_n(z)$ has zeros.

i.e., every domain G_ρ is *simply connected* (cf. footnote 1, p. 258). Obviously, the boundary of E_ρ consists of at most n continua, namely, the boundaries of these domains.[5]

Clearly, if ζ is an exterior point of E_ρ (see p. 65), then $|P_n(\zeta)| > \rho$. In fact, if $|P_n(\zeta)| = \rho$,[6] then $|P_n(z)|$ cannot have a minimum at ζ, since $P_n(\zeta) \neq 0$ (see Theorem 17.5, Corollary). Therefore, in any neighborhood of ζ, we can find points z such that $|P_n(z)| < |P(\zeta)| = \rho$, i.e., points of E_ρ. But this is impossible, since ζ is an exterior point of E_ρ. It follows that $|P_n(z)| < \rho$ on (the interior of) E_ρ, $|P_n(z)| = \rho$ on the boundary of E_ρ, and $|P_n(z)| > \rho$ on the exterior of E_ρ. The boundary of E_ρ, i.e., the locus of all z for which $|P_n(z)| = \rho$, is called a *lemniscate*. If we denote the zeros of $P_n(z)$ by $z_1, \ldots, z_n$ (writing each multiple zero a number of times equal to its order), we find that

$$P_n(z) = a_n(z - z_1) \cdots (z - z_n)$$

[cf. (9.2)], and the equation of the lemniscate becomes

$$|z - z_1| \cdots |z - z_n| = \frac{\rho}{|a_n|} = r^n. \tag{17.11}$$

The points $z_1, \ldots, z_n$ are then called the *foci* of the lemniscate, and the number r defined by (17.11) is called the *radius* of the lemniscate.

If $n = 1$, (17.11) reduces to

$$|z - z_1| = r,$$

which is obviously the equation of a circle of radius r with center at the single focus z_1. If $n = 2$ and $z_1 \neq z_2$, (17.11) becomes

$$|z - z_1|\,|z - z_2| = r^2, \tag{17.12}$$

corresponding to a lemniscate with two foci, whose appearance depends on

$$\mu = \frac{r}{|z_1 - z_2|},$$

the ratio of the radius to the distance between the foci. In Figure 17.1 we draw this lemniscate for the following values of μ:

a) $\mu < \frac{1}{2}$; b) $\mu = \frac{1}{2}$; c) $\frac{1}{2} < \mu < \frac{1}{2}\sqrt{2}$; d) $\frac{1}{2}\sqrt{2} < \mu$.

The lemniscate with equation (17.12) is often called the *oval of Cassini*, or, in the case $\mu = \frac{1}{2}$, the *lemniscate of Bernoulli*.

---

[5] The boundaries of two or more disjoint domains may "coalesce" into a single continuum (see Figures 17.1 and 17.2).

[6] The possibility $|P_n(\zeta)| < \rho$ is immediately precluded, since then ζ would belong to E_ρ.

Remark. For a larger number of foci, we obtain lemniscates of more varied appearance. To show the diversity of these lemniscates we cite without proof the following "lemniscate approximation theorem," due to Hilbert: *The boundary* Γ *of any simply connected, bounded domain* G *can be arbitrarily well approximated by a lemniscate, in the sense that given any* $\varepsilon > 0$, *a lemniscate* Λ *can always be found such that* $\mathcal{N}(z; \varepsilon) \cap \Lambda \neq 0$ *for all* $z \in \Gamma$ *and* $\mathcal{N}(z; \varepsilon) \cap \Gamma \neq 0$ *for all* $z \in \Lambda$.

To see how the shape of a lemniscate with n foci varies as a function of its radius r, consider the lemniscate

$$|P_n(z)| = |z - z_1| \cdots |z - z_n| = r^n, \tag{17.13}$$

where, for simplicity, we set $a_n = 1$ and assume that all the foci $z_1, \ldots, z_n$ are distinct. If $\delta > 0$ is so small that the disks $K_j: |z - z_j| < \delta$ $(j = 1, \ldots, n)$ are disjoint, then

$$E_{r^n} \subset K_1 \cup \cdots \cup K_n$$

for all $r < \delta$ (why?). Moreover, E_{r^n} cannot consist of more than n disjoint domains, and E_{r^n} must contain the center of each disk K_j. Therefore E_{r^n} consists of precisely n disjoint (simply connected) domains, one in each disk K_j, and the lemniscate $|P_n(z)| = r^n$ consists of the n disjoint continua which are the boundaries of these domains.

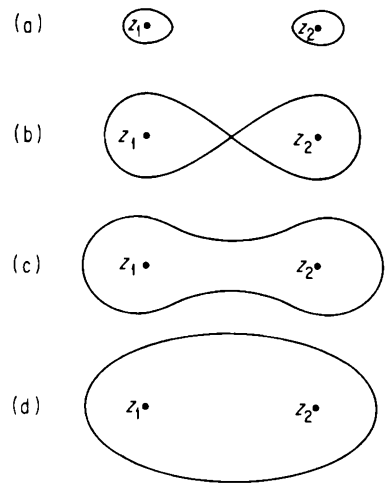

FIGURE 17.1

Now let K be a disk of diameter d containing all the foci of the lemniscate (17.13). Then $|z - z_j| < d$ for all $z \in K$ and $j = 1, \ldots, n$, and hence $|P_n(z)| < d^n$ if $z \in K$, i.e., $K \subset E_r$ if $r > d^n$. Therefore, in this case, E_r consists of just one domain containing K (recall that if E_r consists of several domains, each must contain at least one of the points $z_1, \ldots, z_n$), and the corresponding lemniscate consists of the single continuum which is the boundary of this domain.

To recapitulate, the lemniscate (17.13) with n foci is disconnected (in fact, consists of n disjoint continua) for sufficiently small values of its radius ($r < \delta$), and connected for sufficiently large values of its radius ($r > d^n$). It should also be noted that the lemniscate $|P_n(z)| = r$ belongs to any set $E_{r'}$ with $r' > r$, and hence all the *components* of the lemniscate $|P_n(z)| = r$, i.e., all the continua of which it consists, are contained in the simply connected domains bounded by the components of the lemniscate $|P_n(z)| = r'$. In this

[7] See e.g., E. Hille, *Analytic Function Theory, Vol. II*, Ginn and Co., Boston (1962), Sec. 16.5.

sense, a lemniscate of a given radius is strictly included in any lemniscate of a larger radius (with the same foci). Thus, the change in shape of a lemniscate as its radius r increases can be regarded as a progressive "swelling" of its components. For small r, the components resemble small ovals; as r increases, these ovals gradually enlarge, and at the same time change their shape. For certain critical values of r, the separate components of the lemniscate "coalesce" two, three, or more at a time (see Problem 17.15), thereby diminishing the total number of components. Then further swelling occurs, accompanied from time to time by a further drop in the total number of components, until finally we obtain a lemniscate consisting of a single

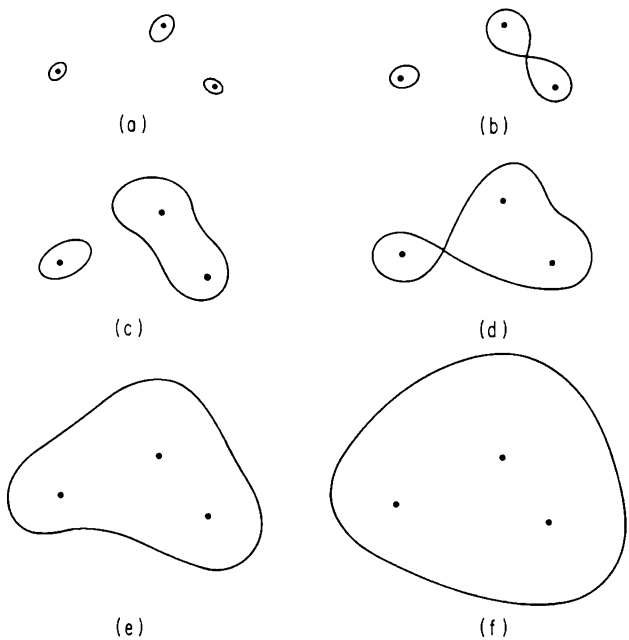

FIGURE 17.2

connected curve. It is lemniscates of this last type that we had in mind in stating Hilbert's lemniscate approximation theorem, and the fact that there is a sufficient variety of such lemniscates to approximate the boundary of any simply connected, bounded domain is due to the great freedom with which one can select the number and location of the foci. In Figure 17.2 we indicate schematically how the shape of a lemniscate varies with the size of r, in the case of three (noncollinear) foci, where the progressive stages of enlargement are consecutively labelled from (a) to (f).

Leaving the subject of lemniscates, we now consider an application of the

maximum modulus principle which plays an important role in problems involving mappings by analytic functions:

THEOREM 17.8 (*Schwarz's lemma*). *Let $f(z)$ be a function which is analytic on the disk $K: |z| < R$ and vanishes at $z = 0$, and suppose that*

$$|f(z)| \leqslant M < \infty \tag{17.14}$$

for all $z \in K$. Then the inequality

$$|f(z)| \leqslant \frac{M}{R}|z| \tag{17.15}$$

holds for all $z \in K$, and moreover

$$|f'(0)| \leqslant \frac{M}{R}. \tag{17.16}$$

Equality is achieved in (17.15) [*for some nonzero $z \in K$*] *or in* (17.16) *if and only if $f(z)$ is an entire linear function of the form*

$$f(z) = \frac{M}{R} e^{i\alpha} z, \tag{17.17}$$

where α is a real number.

Proof. From the Taylor series expansion

$$f(z) = f'(0)\, z + \frac{f''(0)}{2!} z^2 + \cdots \qquad (z \in K)$$

[$f(0) = 0$ by hypothesis], it is apparent that the function

$$\varphi(z) = \frac{f(z)}{z} = f'(0) + \frac{f''(0)}{2!} z + \cdots$$

is analytic on the disk K and takes the value $f'(0)$ at the origin. Let γ_ρ be the circle $|z| = \rho$, where $0 < \rho < R$. Since the inequality

$$|\varphi(z)| = \frac{|f(z)|}{|z|} \leqslant \frac{M}{\rho}$$

holds for all $z \in \gamma_\rho$, it follows from the maximum modulus principle (Theorem 17.5) that

$$|\varphi(z)| \leqslant \frac{M}{\rho}$$

for all $z \in I(\gamma_\rho)$. Therefore, letting $\rho \to R$, we find that

$$|\varphi(z)| \leqslant \frac{M}{R} \tag{17.18}$$

for all $z \in K$. If $z \neq 0$, we substitute for $\varphi(z)$ in (17.18), obtaining

$$|f(z)| \leqslant \frac{M}{R}|z|.$$

This is the desired inequality (17.15), which is obviously valid for $z = 0$. Moreover $\varphi(0) \quad f'(0)$, and hence, when $z = 0$, (17.18) takes the form

$$|f'(0)| \leqslant \frac{M}{R},$$

which is just the inequality (17.16). Finally, we note that (17.15) can become an equality for some nonzero $z \in K$ or (17.16) can become an equality if and only if (17.18) becomes an equality for some $z \in K$. But according to the maximum modulus principle, this is only possible if $\varphi(z) = \text{const}$ on K. Since this constant must have absolute value M/R, it follows that

$$\varphi(z) = \frac{M}{R} e^{i\alpha},$$

or equivalently that

$$f(z) = \frac{M}{R} e^{i\alpha} z.$$

Remark 1. Note that if equality is achieved in (17.15) for a single nonzero $z \in K$, it is achieved for all $z \in K$.

Remark 2. Schwarz's lemma has the following very simple geometric interpretation: Suppose $f(z)$ is analytic on the disk $K: |z| < R$, and suppose the image $f(K)$ of this disk under the mapping $w = f(z)$ is contained in the disk $|w| < M$, where the point $z = 0$ goes into the point $w = 0$. Then the image of any closed disk $|z| \leqslant r < R$, whose radius is $\lambda = r/R$ times smaller than the radius of K, lies in the closed disk $|w| \leqslant (r/R)M$, whose radius is λ times smaller than M. Moreover, the image of any point z_0 lying on the circle $|z| = r$ can only lie on the circle $|w| = (r/R)M$ if the mapping has the form (17.17), i.e., if the mapping consists of a rotation around the origin together with shrinking by a factor M/R relative to the origin.

84. Circular Elements. Regular and Singular Points

By a *circular element of an analytic function* (briefly, a *circular element* or an *element*), we mean a set $\{K, f(z)\}$ consisting of a disk $K: |z - z_0| < \rho$ and an analytic function $f(z)$ defined on K. Given an element $\{K, f(z)\}$, we can divide the points of the circle $\Gamma: |z - z_0| = \rho$ (the boundary of K) into the following two classes (*a priori*, either class can be empty):

1. Points $\zeta \in \Gamma$ for which there can be found a neighborhood $\mathcal{N}(\zeta)$ and an analytic function $\varphi_\zeta(z)$ defined on $\mathcal{N}(\zeta)$, i.e., an element $\{\mathcal{N}(\zeta), \varphi_\zeta(z)\}$, such that $\varphi_\zeta(z) = f(z)$ for all z in $\mathcal{N}(\zeta) \cap K$; such points are called *regular points* [of $\{K, f(z)\}$, or simply of $f(z)$].
2. Points $\zeta \in \Gamma$ for which no such neighborhood and analytic function $\varphi_\zeta(z)$ can be found; such points are called *singular points* [of $\{K, f(z)\}$, or simply of $f(z)$].

It follows from the definition of a regular point that if $\zeta_0 \in \Gamma$ is a regular point, then all the points in $\mathcal{N}(\zeta_0) \cap \Gamma$ are also regular points.[8] In fact, for $\mathcal{N}(\zeta)$ we need only choose any disk with center ζ which is contained in $\mathcal{N}(\zeta_0)$, while for $\varphi_\zeta(z)$ we need only choose the function $\varphi_{\zeta_0}(z)$ which is analytic on $\mathcal{N}(\zeta_0)$ and coincides with $f(z)$ on $\mathcal{N}(\zeta_0) \cap K$. In particular, this means that if $f(z)$ has one regular point, then $f(z)$ has infinitely many regular points. On the other hand, $f(z)$ can have just one singular point.

It should be noted that every point $z_0 \in K$ has the characteristic property of a regular point, i.e., there exists a neighborhood $\mathcal{N}(z_0)$ [in fact, any neighborhood of z_0 contained in K] and an analytic function on $\mathcal{N}(z_0)$ [in fact, $f(z)$ itself] which coincides with $f(z)$ on $\mathcal{N}(z_0) \cap K$. Therefore, extending our definition of a regular point, we shall henceforth regard the points of the disk K itself as regular points of $f(z)$.

Example. Let K be the unit disk $|z| < 1$, and let

$$f(z) = \frac{1}{1 - z}.$$

For every point $\zeta \neq 1$ lying on the boundary of K (i.e., on the unit circle $|z| = 1$), we can find a neighborhood $\mathcal{N}(\zeta): |z - \zeta| < |1 - \zeta|$ and an analytic function

$$\varphi_\zeta(z) = \frac{1}{1 - z}$$

defined on $\mathcal{N}(\zeta)$ such that $\varphi_\zeta(z) = f(z)$ on $\mathcal{N}(\zeta) \cap K$. Therefore every point $\zeta \neq 1$ is a regular point of $f(z)$. On the other hand, the point $\zeta = 1$ is a singular point of $f(z)$, since otherwise there would be a neighborhood $\mathcal{N}(1)$ and an analytic function $\varphi(z)$ coinciding with $f(z)$ on $E = \mathcal{N}(1) \cap K$. But then the limit

$$\varphi(1) = \lim_{\substack{z \to 1 \\ z \in E}} \varphi(z) = \lim_{\substack{z \to 1 \\ z \in E}} \frac{1}{1 - z}$$

would have to be finite, which is impossible.

THEOREM 17.9. *Given a disk $K: |z - z_0| < \rho$ and an element $\{K, f(z)\}$, suppose every point of the circle $\Gamma: |z - z_0| = \rho$ is a regular point. Then there is a disk $K_1: |z - z_0| < \rho_1$ and another element $\{K_1, \varphi(z)\}$, such that $\rho_1 > \rho$ and $\varphi(z) = f(z)$ on K.*

Proof. Let $\mathcal{N}(\zeta)$ be any neighborhood figuring in the definition of the regular point ζ, and let D be the union of K and all the neighborhoods $\mathcal{N}(\zeta)$, $\zeta \in \Gamma$ (see Figure 17.3). Since K and every $\mathcal{N}(\zeta)$ are open, so is D

[8] In a context like this, $\mathcal{N}(\zeta)$ will always denote one of the neighborhoods figuring in the definition of the regular point ζ, i.e., a neighborhood $\mathcal{N}(\zeta)$ in which there exists an analytic function $\varphi_\zeta(z)$ coinciding with $f(z)$ on $\mathcal{N}(\zeta) \cap K$.

(see Problem 3.21). Moreover, D is connected, since any two points z' and z'' in D can be joined by a continuous curve γ contained in D. To see this, we examine the three possible cases in turn:

1. If z' and z'' both belong to the disk K or to the same neighborhood $\mathcal{N}(\zeta)$, we choose γ to be the straight line segment $\overline{z'z''}$; then $\gamma \subset D$, since $\gamma \subset K$ or $\gamma \subset \mathcal{N}(\zeta)$.
2. If one of the points, say z', belongs to K, while the other point z'' belongs to $\mathcal{N}(\zeta)$, we choose γ to be the polygonal curve $\overline{z'\zeta} + \overline{\zeta z''}$; then $\gamma \subset K \cup \mathcal{N}(\zeta) \subset D$, since $\overline{z'\zeta} - \{\zeta\} \subset K$ and $\overline{\zeta z''} \subset \mathcal{N}(\zeta)$.
3. Finally, if $z' \in \mathcal{N}(\zeta')$, $z'' \in \mathcal{N}(\zeta'')$, we choose γ to be the curve $\overline{z'\zeta'} + \overset{\frown}{\zeta'\zeta''} + \overline{\zeta''z''}$, where $\overset{\frown}{\zeta'\zeta''}$ is either arc of Γ joining ζ' to ζ'' (see Figure 17.3); then $\gamma \subset D$, since $\overline{z'\zeta'} \in \mathcal{N}(\zeta')$, $\overline{\zeta''z''} \in \mathcal{N}(\zeta'')$, while $\overset{\frown}{\zeta'\zeta''}$ consists entirely of points which are centers of neighborhoods $\mathcal{N}(\zeta)$, $\zeta \in \Gamma$.

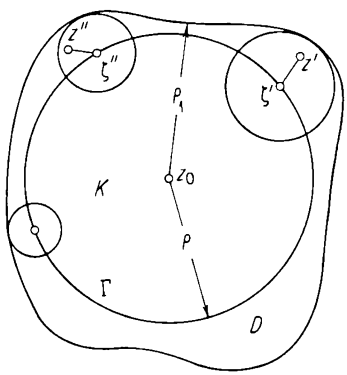

FIGURE 17.3

Thus D is an open connected set, i.e., a domain.

Next we define a function $\varphi(z)$ on D by setting

$$\varphi(z) = \begin{cases} f(z) & \text{if } z \in K, \\ \varphi_\zeta(z) & \text{if } z \in \mathcal{N}(\zeta), \end{cases} \tag{17.19}$$

where $\varphi_\zeta(z)$ has the same meaning as in the definition of the regular point ζ. As we now show, $\varphi(z)$ is single-valued and analytic on D. Since D contains points which simultaneously belong to a neighborhood $\mathcal{N}(\zeta)$ and to K, and also points which simultaneously belong to two neighborhoods $\mathcal{N}(\zeta')$ and $\mathcal{N}(\zeta'')$, we must verify that the definition (17.19) is *consistent*, in the sense that a point $z \in D$ will always be assigned the same value $\varphi(z)$ whether z is regarded as belonging to K, $\mathcal{N}(\zeta)$, $\mathcal{N}(\zeta')$ or $\mathcal{N}(\zeta'')$. There are basically two cases to be considered:

1. Let $G = \mathcal{N}(\zeta) \cap K$. Then $\varphi_\zeta(z) = f(z)$ on G, by the very definition of the regular point ζ.
2. Let $G = \mathcal{N}(\zeta') \cap \mathcal{N}(\zeta'') \neq 0$ and let $E = G \cap K$ (see Figure 17.4, which illustrates the situation where neither neighborhood $\mathcal{N}(\zeta')$, $\mathcal{N}(\zeta'')$ is a subset of the other). Then $\varphi_{\zeta'}(z) = f(z) = \varphi_{\zeta''}(z)$ on

E, and hence, by the interior uniqueness theorem, $\varphi_{\zeta'}(z) = \varphi_{\zeta''}(z)$ on the whole set G, since $E \subset G$ obviously contains a limit point of G.

Thus the function $\varphi(z)$ defined by (17.9) is single-valued on D. Moreover, $\varphi(z)$ is analytic on D, since, by construction, $\varphi(z)$ coincides with an analytic function [either $f(z)$ or some $\varphi_\zeta(z)$] in a neighborhood of every point of D [note that if $z \in D$, then there is a neighborhood $\mathcal{N}(z)$ such that $\mathcal{N}(z) \subset K$ or $\mathcal{N}(z) \subset \mathcal{N}(\zeta)$ for some $\zeta \in \Gamma$].

Now let D^b be the boundary of D, and let ρ_1 be the distance between z_0 and D^b. Then ρ_1 is the distance between z_0 and some point $z_1 \in D^b$ (see Theorem 3.20).[9] But z_1 belongs to the exterior of the circle Γ (note that $\Gamma \subset D$ while $z_1 \in D^b$), and hence ρ_1 exceeds the radius of Γ, i.e., $\rho_1 > \rho$. It is clear that the disk $K_1: |z - z_0| < \rho_1$ and the function $\varphi(z)$ satisfy the requirements of the theorem, since $\varphi(z)$ is analytic on K_1 and coincides with $f(z)$ on K.

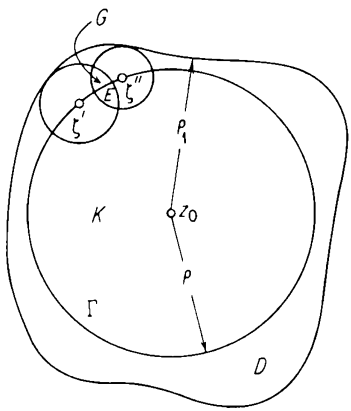

FIGURE 17.4

THEOREM 17.10. *If the element* $\{K, f(z)\}$ *has a singular point* ζ *on the boundary of the disk* $K: |z - z_0| < \rho$, *then* R, *the radius of convergence of the Taylor series*

$$f(z_0) + f'(z_0)(z - z_0) + \cdots + \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n + \cdots, \quad (17.20)$$

equals ρ, *but otherwise* R *exceeds* ρ.

Proof. Let $\Gamma: |z - z_0| = \rho$ be the boundary of K, and suppose Γ contains at least one singular point ζ. According to Theorem 16.7, $R \geqslant \rho$. Suppose $R \neq \rho$, which implies $R > \rho$. The series (17.20) defines an analytic function on the disk $|z - z_0| < R$, in particular, in any neighborhood $\mathcal{N}(\zeta)$ contained in this disk (such neighborhoods exist, since $|\zeta| = \rho < R$). Moreover, the analytic function defined by (17.20) obviously coincides with $f(z)$ on $\mathcal{N}(\zeta) \cap K$. But this is impossible, since ζ is a singular point of $\{K, f(z)\}$, and hence $R = \rho$.

If there are no singular points on Γ, then every point of Γ is a regular point. It follows from Theorem 17.9 that there is a disk $K_1: |z - z_0| < \rho_1$ and another element $\{K_1, \varphi(z)\}$ such that $\rho_1 > \rho$ and $\varphi(z) = f(z)$ on K.

---

[9] There is no loss of generality in assuming that the radii of the neighborhoods $\mathcal{N}(\zeta)$, $\zeta \in \Gamma$ have a common upper bound, so that D is bounded and D^b is compact.

But $\varphi(z)$ and $f(z)$ have the same Taylor series expansion (17.20) at the point z_0, and hence $R \geqslant \rho_1 > \rho$.

Before proceeding to the next result, we note that if the radius of convergence of the power series

$$f(z) = a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots \tag{17.21}$$

equals R, then the disk $K: |z - z_0| < R$ and the sum function $f(z)$ constitute a circular element $\{K, f(z)\}$. For simplicity, we shall refer to a regular or singular point of the element $\{K, f(z)\}$ as a regular or singular point of the series (17.21), or of its sum $f(z)$.

THEOREM 17.11. *The power series* (17.21) *has at least one singular point on its circle of convergence* $\Gamma: |z - z_0| = R$.

Proof. Let K be the interior of Γ. Then $\{K, f(z)\}$ is an element. Suppose $\{K, f(z)\}$ has no singular points on Γ. Then, since the Taylor series of the sum of a power series is obviously the power series itself, it follows from Theorem 17.10 that the radius of convergence of (17.21) exceeds R, which is absurd. Therefore $\{K, f(z)\}$ has at least one singular point on Γ, and the theorem is proved.

Example 1. The geometric series

$$1 + z + z^2 + \cdots + z^n + \cdots$$

has circle of convergence $\Gamma: |z| = 1$ and sum $1/(1 - z)$. As required by Theorem 17.11, there is a singular point on Γ, namely, the point $z = 1$ (see the example on p. 383).

Example 2. We now show that *every* point on the circle of convergence $\Gamma: |z| = 1$ of the power series

$$f(z) = 1 + z^2 + z^4 + \cdots + z^{2^n} + \cdots$$

is a singular point. Suppose first that $z \to 1$ along the segment $\overline{01}$, i.e., along the radius of Γ contained in the nonnegative real axis. Then, since the nth partial sum

$$1 + x^2 + x^4 + \cdots + x^{2^n}$$

approaches $n + 1$ as $x \to 1-$, there is a $\delta(n) > 0$ such that

$$1 + x^2 + x^4 + \cdots + x^{2^n} > n$$

if $x > 1 - \delta(n)$. Therefore

$$f(x) = \sum_{k=0}^{\infty} x^{2^k} > \sum_{k=0}^{n} x^{2^k} > n$$

if $x > 1 - \delta(n)$, which implies

$$\lim_{x \to 1-} f(x) = \infty.$$

Using this fact, we can verify that $z = 1$ is a singular point of $f(z)$, by the same argument as in the case of the geometric series.

Next we note the identity

$$f(z) = z^2 + z^4 + \cdots + z^{2^n} + [1 + (z^{2^n})^2 + (z^{2^n})^4 + \cdots].$$

Since the term in brackets differs from the original series only by having z^{2^n} instead of z, we have

$$f(z) = z^2 + z^4 + \cdots + z^{2^n} + f(z^{2^n})$$

for any positive integer n. Consider all the 2^nth roots of unity, i.e., all the numbers

$$\sqrt[2^n]{1}, \tag{17.22}$$

which correspond to the vertices of a regular 2^n-gon inscribed in the unit circle. If ζ is one of these roots and if $z \to \zeta$ along the radius $\overline{0\zeta}$, then z^{2^n} obviously lies on the radius $\overline{01}$ and $z^{2^n} \to 1-$ as $z \to \zeta$. It follows that

$$\lim_{\substack{z \to \zeta \\ z \in \overline{0\zeta}}} f(z^{2^n}) = \infty,$$

and therefore

$$\lim_{\substack{z \to \zeta \\ z \in \overline{0\zeta}}} f(z) = \lim_{\substack{z \to \zeta \\ z \in \overline{0\zeta}}} [z^2 + z^4 + \cdots + z^{2^n} + f(z^{2^n})] = \infty.$$

Thus every root (17.22) is a singular point of $f(z)$, for all $n = 1, 2, \ldots$. Consequently, the set of singular points of $f(z)$ is *everywhere dense* in the unit circle Γ (i.e., every arc of Γ, however small, contains a singular point). But this means that *every* point of Γ is a singular point, since if $z_1 \in \Gamma$ were a regular point, then, as on p. 383, some arc $\mathcal{N}(z_0) \cap \Gamma$ would consist only of regular points, which is impossible.

The next theorem gives a general method for deciding whether or not a point on the circle of convergence of a power series is a regular point or a singular point:

THEOREM 17.12. *Given a power series*

$$f(z) = a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots, \tag{17.23}$$

with radius of convergence R, let ζ be a point of the circle $\Gamma: |z - z_0| = R$ and let z_1 be a point of the radius $\overline{z_0\zeta}$ which differs from z_0 and ζ. Then ζ is a singular point of $f(z)$ if

$$\Delta = R - |z_1 - z_0| = \frac{1}{\overline{\lim}_{n \to \infty} \sqrt[n]{\dfrac{|f^{(n)}(z_1)|}{n!}}}, \tag{17.24}$$

and a regular point of $f(z)$ if

$$\Delta < \frac{1}{\overline{\lim}_{n\to\infty} \sqrt[n]{\dfrac{|f^{(n)}(z_1)|}{n!}}}. \tag{17.25}$$

Proof. Consider the series

$$\tilde{f}(z) = b_0 + b_1(z - z_1) + \cdots + b_n(z - z_1)^n + \cdots, \tag{17.26}$$

where

$$b_n = \frac{f^{(n)}(z_1)}{n!} = a_n + \frac{n+1}{1!}a_{n+1}(z_1 - z_0) + \frac{(n+1)(n+2)}{2!}a_{n+2}(z_1 - z_0)^2 + \cdots \qquad (n = 0, 1, 2, \ldots).$$

According to the Cauchy-Hadamard formula (16.2), the radius of convergence of the series (17.26) is

$$r = \frac{1}{\overline{\lim}_{n\to\infty} \sqrt[n]{|b_n|}}.$$

On the other hand, (17.26) is the Taylor series expansion of $f(z)$ in powers of $z - z_1$ (note the definition of the coefficients b_n), and therefore, according to Theorem 16.7, (17.26) certainly converges on the disk $|z - z_1| < \Delta$, where $\Delta = R - |z_1 - z_0|$ is the distance between z_1 and Γ. It follows that $r \geqslant \Delta$.

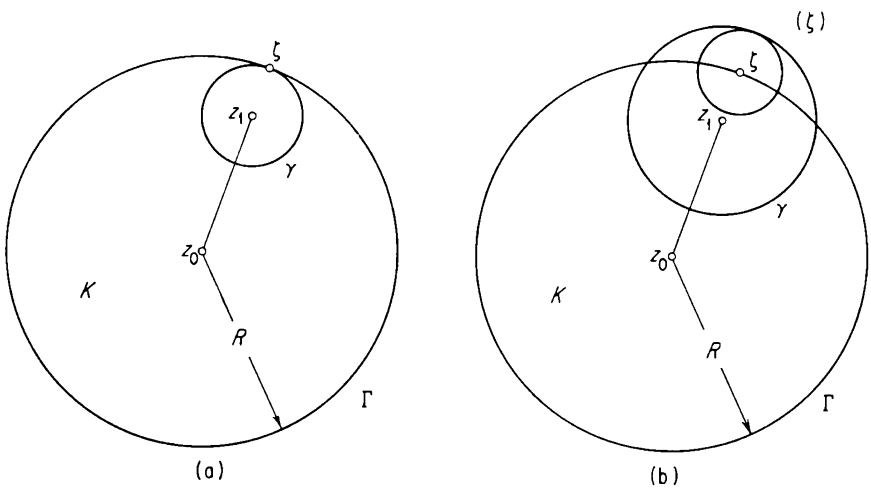

FIGURE 17.5

First suppose that $r = \Delta$, as shown in Figure 17.5(a). Then, according to Theorem 17.11, the circle $\gamma: |z - z_1| = r$ contains at least one singular point of $\tilde{f}(z)$. But no point $\zeta' \in \gamma \cap K$, where $K = I(\Gamma)$, can be a singular point of $\tilde{f}(z)$, since every such ζ' has a neighborhood $\mathcal{N}(\zeta') \subset K$ in which $f(z)$, i.e., the sum of the original series (17.23), is an analytic function coinciding with $\tilde{f}(z)$ on $\mathcal{N}(\zeta') \cap I(\gamma)$. Therefore the point ζ is the unique singular point of $\tilde{f}(z)$. Moreover, ζ is also a singular point of $f(z)$, since otherwise there would be a function $\varphi(z)$ analytic in some neighborhood $\mathcal{N}(\zeta)$ and coinciding with $f(z)$ on $\mathcal{N}(\zeta) \cap K$. But then $\varphi(z)$ would coincide with $\tilde{f}(z)$ on $\mathcal{N}(\zeta) \cap I(\gamma)$, since $f(z) = \tilde{f}(z)$ on $\mathcal{N}(\zeta) \cap I(\gamma)$, contradicting the fact that ζ is a singular point of $\tilde{f}(z)$. This proves that ζ is a singular point of $f(z)$ if (17.24) holds.

Next suppose that $r \neq \Delta$, i.e., $r > \Delta$. Then the circle $\gamma: |z - z_0| = r$ *intersects* Γ, as shown in Figure 17.5(b). Therefore, there is a neighborhood $\mathcal{N}(\zeta) \subset I(\gamma)$ in which $\tilde{f}(z)$ represents an analytic function coinciding with $f(z)$ on $\mathcal{N}(\zeta) \cap K$. In other words, ζ is a regular point of $f(z)$ if (17.25) holds, and the theorem is proved.

As an application of Theorem 17.12, we have the following result:

THEOREM 17.13 (*Pringsheim's theorem*). *Given a power series*

$$f(z) = a_0 + a_1 z + \cdots + a_n z^n + \cdots, \tag{17.27}$$

with radius of convergence 1, *suppose the coefficients* a_n *are all nonnegative real numbers. Then* $z = 1$ *is a singular point of* $f(z)$.

Proof. Suppose $z = 1$ is not a singular point of $f(z)$. Then, choosing $R = 1$, $z_0 = 0$ and $z_1 = x$ $(0 < x < 1)$ in (17.23), we have

$$\Delta = 1 - x < \frac{1}{\overline{\lim}_{n \to \infty} \sqrt[n]{\dfrac{|f^{(n)}(x)|}{n!}}}. \tag{17.28}$$

If ζ is any other point of the unit circle $\Gamma: |z| = 1$, let z_1 be the point in which the radius $\overline{0\zeta}$ intersects the circle $|z| = x$, so that $|z_1| = x$. Then the distance from z_1 to ζ is again $\Delta = 1 - x$. On the other hand,

$$\begin{aligned} \frac{|f^{(n)}(z_1)|}{n!} &= \left| a_n + \frac{n+1}{1!} a_{n+1} z_1 + \frac{(n+1)(n+2)}{2!} a_{n+2} z_1^2 + \cdots \right| \\ &\leqslant a_n + \frac{n+1}{1!} a_{n+1} x + \frac{(n+1)(n+2)}{2!} a_{n+2} x^2 + \cdots \\ &= \frac{|f^{(n)}(x)|}{n!}, \end{aligned}$$

and hence

$$\frac{1}{\overline{\lim}_{n \to \infty} \sqrt[n]{\dfrac{|f^{(n)}(z_1)|}{n!}}} \geqslant \frac{1}{\overline{\lim}_{n \to \infty} \sqrt[n]{\dfrac{|f^{(n)}(x)|}{n!}}}. \tag{17.29}$$

It follows from (17.28) and (17.29) that

$$\Delta = 1 - |z_1| < \frac{1}{\overline{\lim}_{n \to \infty} \sqrt[n]{\dfrac{|f^{(n)}(z_1)|}{n!}}},$$

i.e., ζ is a regular point of $f(z)$. Since ζ is arbitrary, every point of Γ is a regular point, contrary to the assumption that Γ is the circle of convergence of (17.27) [cf. Theorem 17.11]. Therefore $z = 1$ must be a singular point of $f(z)$, and the theorem is proved.

Remark. Instead of the point 1, consider any other point ζ of the unit circle. Then exactly the same argument shows that ζ is a singular point of the series (17.27) if the numbers $a_n\zeta^n$ $(n = 0, 1, 2, \ldots)$ are real and nonnegative. In fact, it is sufficient to require only that these numbers be real and nonnegative for all n exceeding a certain integer n_0, since writing $f(z)$ in the form

$$f(z) = \sum_{n=0}^{n_0} a_n z^n + \sum_{n=n_0+1}^{\infty} a_n z^n,$$

we see at once that ζ is a singular point of $f(z)$ if and only if ζ is a singular point of the second series on the right.

Example. According to the Cauchy-Hadamard formula, the series

$$f(z) = \sum_{n=0}^{\infty} \frac{z^{2^n}}{2^{n^2}} \tag{17.30}$$

has radius of convergence 1, since

$$\overline{\lim_{n \to \infty}} \sqrt[n]{|a_n|} = \lim_{n \to \infty} \sqrt[2^n]{\frac{1}{2^{n^2}}} = 1.$$

It follows from Theorem 17.13 that the point $z = 1$ is a singular point of (17.30). Moreover, according to the above remark, every point $\zeta = \sqrt[2^k]{1}$, where k is an arbitrary positive integer, is also a singular point of (17.30), since if $n \geqslant k$,

$$\frac{\zeta^{2^n}}{2^{n^2}} = \frac{(\zeta^{2^k})^{2^{n-k}}}{2^{n^2}} = \frac{1}{2^{n^2}} > 0.$$

Thus the set of singular points of (17.30) is everywhere dense in the unit circle $\Gamma: |z| = 1$. But then, by the argument given on p. 387, *every* point of Γ is a singular point of (17.30).

Remarkably enough, the fact that every point of Γ is a singular point does not prevent the series (17.30) from converging absolutely and uniformly

on the closed unit disk $\bar{K}: |z| \leqslant 1$, and, in fact, its sum $f(z)$ is *infinitely differentiable* on $\bar{K}$. To see this, we first note that

$$\left|\frac{z^{2^n}}{2^{n^2}}\right| \leqslant \frac{1}{2^{n^2}}$$

if $|z| \leqslant 1$. Therefore, since the series

$$\sum_{n=0}^{\infty} \frac{1}{2^{n^2}}$$

is convergent, the series (17.30) is absolutely and uniformly convergent on $\bar{K}$ (cf. Theorem 15.2). It follows from Theorem 15.3 that $f(z)$ is continuous on $\bar{K}$. Moreover, differentiating the series (17.30) term by term k times, we obtain the series[10]

$$\sum_{n=0}^{\infty} \frac{2^n(2^n - 1) \cdots (2^n - k + 1)}{2^{n^2}} z^{2^n - k}, \tag{17.31}$$

whose terms have absolute values satisfying the inequality

$$\left|\frac{2^n(2^n - 1) \cdots (2^n - k + 1)}{2^{n^2}}\right| \leqslant \frac{2^{nk}}{2^{n^2}} = \frac{1}{2^{n(n-k)}} \leqslant \frac{1}{2^n}$$

if $n \geqslant k$ and $|z| \leqslant 1$. Therefore the series (17.31) is uniformly convergent on $\bar{K}$, and hence represents the derivative $f^{(k)}(z)$ on $\bar{K}$ (see Problem 17.19). Thus $f(z)$ is a continuous, infinitely differentiable function on $\bar{K}$, which is analytic on the open disk $K: |z| < 1$, although every point of Γ is a singular point of $f(z)$, as already noted. The presence of these singular points has no "outward effect" on the behavior of the function on Γ, at least not to the extent of destroying the continuity of $f(z)$ or of its derivatives on Γ. In fact, to show that every point $\zeta \in \Gamma$ is singular, we need only examine the derivatives of $f(z)$ at *interior* points $z_1 \in K$, by comparing the quantity

$$\frac{1}{\overline{\lim}_{n \to \infty} \sqrt[n]{\dfrac{f^{(n)}(z_1)}{n!}}}$$

with the distance from z_1 to ζ.

Given a single-valued analytic function $f(z)$ defined on a domain G, the problem of finding Taylor series expansions of $f(z)$ at various points of G reduces to a study of the (infinitely many) circular elements $\{K, f(z)\}$, where $K \subset G$. Because of Theorem 17.10, the location of the singular points of these elements plays a key role. It should be noted that

1. Any singular point of an element $\{K, f(z)\}$, $K \subset G$ must belong to Γ, the boundary of G (why?).

---

[10] The coefficient of $z^{2^n - k}$ in (17.31) is taken to be 0 if $2^n < k$.

2. There may be points of Γ which are singular points of some elements $\{K, f(z)\}$, $K \subset G$, and regular points of others (see Example 2 below).

Example 1. The function

$$f(z) = \frac{1}{1 + z^2}$$

is single-valued and analytic in the whole plane, with the exception of the points $\pm i$ at which $f(z)$ becomes infinite. Let z_0 be an arbitrary point different from $\pm i$, and draw the circle $\gamma: |z - z_0| = \rho$ passing through whichever point i or $-i$ is nearer to z_0, say i. Clearly, the point i is a singular point of the element $\{I(\gamma), f(z)\}$, since otherwise there would be a neighborhood $\mathcal{N}(i)$ and an analytic function $\varphi(z)$ coinciding with $f(z)$ on $E = \mathcal{N}(i) \cap I(\gamma)$. But then the limit

$$\varphi(i) = \lim_{\substack{z \to i \\ z \in E}} \varphi(z) = \lim_{\substack{z \to i \\ z \in E}} \frac{1}{1 + z^2}$$

would have to be finite, which is impossible (cf. the example on p. 383).[11] Therefore, according to Theorem 17.10, the circle of convergence of the Taylor series

$$f(z_0) + f'(z_0)(z - z_0) + \cdots + \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n + \cdots \tag{17.32}$$

is precisely γ. In other words, the radius of convergence of (17.32) is just ρ, the distance from z_0 to the nearer of the two points i and $-i$.

The simplest way of finding the Taylor series of $f(z)$ is to expand $(1 + z^2)^{-1}$ in partial fractions and then express the result in terms of geometric series. Thus we write

$$\begin{aligned} \frac{1}{1 + z^2} &= -\frac{1}{2i}\left(\frac{1}{i - z} + \frac{1}{i + z}\right) \\ &= -\frac{1}{2i}\left(\frac{1}{i - z_0}\frac{1}{1 - \dfrac{z - z_0}{i - z_0}} + \frac{1}{i + z_0}\frac{1}{1 + \dfrac{z - z_0}{i + z_0}}\right), \end{aligned}$$

where

$$\left|\frac{z - z_0}{i - z_0}\right| < 1, \qquad \left|\frac{z - z_0}{i + z_0}\right| < 1,$$

since $z \in I(\gamma)$ but $\pm i \notin I(\gamma)$. It follows that

$$\begin{aligned} \frac{1}{1 + z^2} &= -\frac{1}{2i}\left[\frac{1}{i - z_0}\sum_{n=0}^{\infty}\left(\frac{z - z_0}{i - z_0}\right)^n + \frac{1}{i + z_0}\sum_{n=0}^{\infty}(-1)^n\left(\frac{z - z_0}{i + z_0}\right)^n\right] \\ &= \frac{1}{2i}\sum_{n=0}^{\infty}(-1)^{n+1}[(z_0 + i)^{-n-1} - (z_0 - i)^{-n-1}](z - z_0)^n \end{aligned} \tag{17.33}$$

[11] This argument is quite general, as shown in Problem 17.20.

for all $|z| < \rho$. If $z_0 = 0$, then $\rho = 1$ and (17.33) becomes[12]

$$\frac{1}{1+z^2} = \frac{1}{2}\sum_{n=0}^{\infty} i^n(1 + i^{2n})z^n = \sum_{n=0}^{\infty} (-1)^n z^{2n},$$

while if $z_0 = 1$, we have $\rho = \sqrt{2}$ and (17.33) becomes

$$\begin{aligned}\frac{1}{1+z^2} &= \sum_{n=0}^{\infty} (-1)^n\, 2^{-(n+1)/2} \sin\left[\tfrac{1}{4}(n+1)\pi\right](z-1)^n \\ &= \tfrac{1}{2} - \tfrac{1}{2}(z-1) + \tfrac{1}{4}(z-1)^2 - \tfrac{1}{8}(z-1)^4 + \cdots.\end{aligned}$$

Example 2. Let G be the domain whose boundary consists of the non-negative real axis. From all the single-valued branches of the multiple-valued function

$$F(z) = \frac{1}{\operatorname{Ln} z} = \frac{1}{\ln|z| + i \operatorname{Arg} z}$$

defined on G, choose the branch

$$f(z) = \frac{1}{\operatorname{Ln}_1 z} = \frac{1}{\ln|z| + i \operatorname{Arg}_1 z},$$

where $\operatorname{Arg}_1 z$ is the value of $\operatorname{Arg} z$ satisfying the inequality

$$0 < \operatorname{Arg}_1 z < 2\pi.$$

Clearly, $f(z)$ is analytic on G (cf. Sec. 56). We now examine some typical elements $\{K, f(z)\}$, $K \subset G$. *A priori*, one might anticipate that some of these elements will have a singular point at $z = 0$, where $\operatorname{Ln} z$ has a branch point, and that others will have a singular point at $z = 1$, where one of the values of $\operatorname{Ln} z$ vanishes.

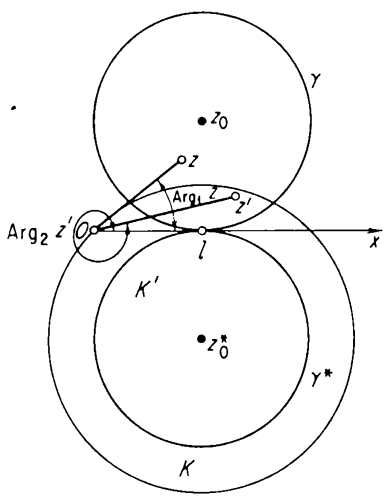

FIGURE 17.6

First let z_0 be the point $1 + i$, lying in the upper half-plane, and draw the circle $\gamma: |z - z_0| = 1$ passing through the point 1, which is the closer of the two points 0, 1 to z_0 (see Figure 17.6). Then the point 1 is a singular point of the element $\{I(\gamma), f(z)\}$. In fact, since

$$\lim_{\substack{z \to 1 \\ z \in I(\gamma)}} \ln|z| = 0, \qquad \lim_{\substack{z \to 1 \\ z \in I(\gamma)}} \operatorname{Arg}_1 z = 0,$$

we have

$$\lim_{\substack{z \to 1 \\ z \in I(\gamma)}} f(z) = \infty,$$

[12] Cf. the discussion of the real power series (1.12).

and therefore, by the usual argument, there is no neighborhood $\mathcal{N}(1)$ and analytic function $\varphi(z)$ defined on $\mathcal{N}(1)$ such that $\varphi(z) = f(z)$ on $\mathcal{N}(1) \cap I(\gamma)$. It follows from Theorem 17.10 that R, the radius of convergence of the Taylor series

$$f(z_0) + f'(z_0)(z - z_0) + \cdots + \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n + \cdots,$$

equals 1.

Next let z_0^* be the point $1 - i$, lying in the lower half-plane, and draw the circle $\gamma^*: |z - z_0^*| = 1$ passing through the point 1, which is again the closer of the two points 0, 1 to z_0^*. Since this time

$$\lim_{\substack{z \to 1 \\ z \in I(\gamma^*)}} \ln |z| = 0, \qquad \lim_{\substack{z \to 1 \\ z \in I(\gamma^*)}} \operatorname{Arg}_1 z = 2\pi \neq 0,$$

it is not true that

$$\lim_{\substack{z \to 1 \\ z \in I(\gamma^*)}} f(z) = \infty,$$

which would imply that 1 is a singular point of the element $\{I(\gamma^*), f(z)\}$. On the contrary, every point of γ^* is a regular point of $\{I(\gamma^*), f(z)\}$. To prove this, consider the disk $K: |z - z_0^*| < \sqrt{2}$ whose boundary contains the point 0 (see Figure 17.6). On K we can define a single-valued analytic branch of $\operatorname{Ln} z$ which coincides with $\operatorname{Ln}_1 z$ on $I(\gamma^*) \subset K$. In fact, if we denote this branch by

$$\operatorname{Ln}_2 z = \ln |z| + i \operatorname{Arg}_2 z,$$

it is clear that $\operatorname{Arg}_2 z$ satisfies the inequality

$$\frac{5\pi}{4} < \operatorname{Arg}_2 z < \frac{9\pi}{4}$$

for all $z \in K$, since $\operatorname{Arg}_2 z_0^*$ must equal

$$\operatorname{Arg}_1 z_0^* = \operatorname{Arg}_1 (1 - i) = \frac{7\pi}{4},$$

while K lies to the right of the line of slope $\pi/4$ passing through the origin. It follows that $\operatorname{Ln}_2 z$ does not vanish at any point of K, and hence the function

$$\varphi(z) = \frac{1}{\operatorname{Ln}_2 z}$$

is analytic on K and coincides with $f(z)$ on $I(\gamma^*)$. Therefore every point $\zeta \in \gamma^*$ is a regular point of the element $\{I(\gamma^*), f(z)\}$, since $\varphi(z)$ is analytic in any neighborhood $\mathcal{N}(\zeta) \subset K$ and coincides with $f(z)$ on $\mathcal{N}(\zeta) \cap I(\gamma^*)$.[13] In particular, the point 1 is a regular point of the element $\{I(\gamma^*), f(z)\}$, despite the fact that it is a singular point of the element $\{I(\gamma), f(z)\}$.

---

[13] Cf. footnote 8, p. 383.

Finally, consider the Taylor series

$$f(z_0^*) + f'(z_0^*)(z - z_0^*) + \cdots + \frac{f^{(n)}(z_0^*)}{n!}(z - z_0^*)^n + \cdots,$$

which is simultaneously the Taylor series of the function $f(z)$ defined on $I(\gamma^*)$ and the function $\varphi(z)$ defined on K. If R' is the radius of convergence of this series, then obviously $R' \geqslant \sqrt{2}$. However, the origin $z = 0$ is a singular point of the element $\{K, \varphi(z)\} = \{K, \operatorname{Ln}_2 z\}$, since

$$\lim_{\substack{z \to 0 \\ z \in K}} \frac{d}{dz} \frac{1}{\operatorname{Ln}_2 z} = \lim_{\substack{z \to 0 \\ z \in K}} \frac{-1}{z(\operatorname{Ln}_2 z)^2} = \infty,$$

and hence there is no neighborhood $\mathcal{N}(0)$ and analytic function $\psi(z)$ defined on $\mathcal{N}(0)$ such that $\psi(z) = \varphi(z)$ on $\mathcal{N}(0) \cap K$ (why not?). It follows from Theorem 17.10 that $R' = \sqrt{2}$.

This example is studied further in Problem 17.21.

*85. Behavior of a Power Series on its Circle of Convergence

A power series can behave in any of the following three ways on its circle of convergence Γ:

1. The series may diverge everywhere on Γ, as shown by the geometric series

$$\sum_{n=0}^{\infty} z^n,$$

which diverges for all $|z| = 1$.

2. The series may converge at some points of Γ and diverge at other points of Γ, as shown by the logarithmic series

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{z^n}{n},$$

which converges for $z = 1$ and diverges for $z = -1$.

3. The series may converge everywhere on Γ, as shown by the example

$$\sum_{n=1}^{\infty} \frac{z^n}{n^2}$$

which converges (absolutely) for all $|z| = 1$.

We now inquire as to what convergence of a power series at a point $\zeta \in \Gamma$ implies about the behavior of the sum of the series in a neighborhood of ζ. First we note that there is no loss of generality in assuming that Γ is the unit

circle $|z| = 1$ and ζ is the point $z = 1$. In fact, the Möbius transformation

$$z = \frac{Z - Z_0}{\zeta - Z_0} \tag{17.34}$$

converts the general power series

$$A_0 + A_1(Z - Z_0) + \cdots + A_n(Z - Z_0)^n + \cdots,$$

with circle of convergence $|Z - Z_0| = R$, into a new power series of the form

$$a_0 + a_1 z + \cdots + a_n z^n + \cdots,$$

with circle of convergence $|z| = 1$, and moreover (17.34) maps the point $Z = \zeta$ into the point $z = 1$.

THEOREM 17.14 (*Abel's limit theorem*). *Suppose the power series*

$$\sum_{n=0}^{\infty} a_n z^n, \tag{17.35}$$

with circle of convergence $\Gamma: |z| = 1$, *converges at the point* $z = 1$. *Let* G_θ *be the interior of any angle* $2\theta < \pi$ *radians with vertex at* $z = 1$, *such that* G_θ *opens to the left and is symmetric with respect to the radius* $\overline{01}$. *Then*

$$\lim_{\substack{z \to 1 \\ z \in G_\theta}} f(z) = \sum_{n=0}^{\infty} a_n$$

where $f(z)$ *is the sum of the series* (17.35).[14]

Proof. Write the difference

$$\Delta(z) = \sum_{k=0}^{\infty} a_k - f(z) = \sum_{k=0}^{\infty} a_k(1 - z^k)$$

in the form

$$\Delta(z) = \sum_{k=0}^{n} a_k(1 - z^k) + \sum_{k=n+1}^{\infty} a_k(1 - z^k), \tag{17.36}$$

where n is a fixed but arbitrary positive integer. Since obviously the first term in the right-hand side of (17.36) approaches 0 as $z \to 1$, for any fixed n, we need only prove that the second term can be made arbitrarily small for $z \in G_0$ and sufficiently large n. With this in mind, consider the sum

$$\sum_{k=n+1}^{n+p} a_k(1 - z^k) = (1 - z) \sum_{k=n+1}^{n+p} a_k(1 + z + \cdots + z^{k-1}),$$

[14] Thus $z \to 1$ along a nontangential path (cf. Lemma 3 and Definition 2, p. 311).

and introduce the notation

$$\sum_{k=n+1}^{m} a_k = \alpha_m \quad (m > n), \qquad \alpha_n = 0,$$

so that

$$a_k = \alpha_k - \alpha_{k-1} \qquad (k = n+1, \quad n+2, \ldots).$$

Moreover, let

$$1 + z + \cdots + z^{k-1} = b_k.$$

Then, making an Abel transformation as in equation (3.16), p. 38, we find that

$$\begin{aligned}\sum_{k=n+1}^{n+p} a_k(1 - z^k) &= (1 - z) \sum_{k=n+1}^{n+p} (\alpha_k - \alpha_{k-1})b_k \\ &= (1 - z)\left[\alpha_{n+p}b_{n+p} - \sum_{k=n+1}^{n+p-1} \alpha_k(b_{k+1} - b_k)\right] \\ &= (1 - z)\left[\sum_{k=n+1}^{n+p} a_k \cdot \sum_{k=0}^{n+p-1} z^k - \sum_{k=n+1}^{n+p-1}\left(\sum_{l=n+1}^{k} a_l\right)z^k\right].\end{aligned} \tag{17.37}$$

Since if $|z| < 1$, the sums

$$\sum_{k=n+1}^{n+p} a_k(1 - z^k) = \sum_{k=n+1}^{n+p} a_k - \sum_{k=n+1}^{n+p} a_k z^k, \quad \sum_{k=n+1}^{n+p} a_k, \quad \sum_{k=0}^{n+p-1} z^k$$

all converge to finite limits as $p \to \infty$ (the corresponding infinite series converge), the same is true of the sum

$$\sum_{k=n+1}^{n+p-1}\left(\sum_{l=n+1}^{k} a_l\right)z^k.$$

Therefore, taking the limit as $p \to \infty$ in (17.37), we obtain

$$\begin{aligned}\sum_{k=n+1}^{\infty} a_k(1 - z^k) &= (1 - z)\left[\sum_{k=n+1}^{\infty} a_k \cdot \sum_{k=0}^{\infty} z^k - \sum_{k=n+1}^{\infty}\left(\sum_{l=n+1}^{k} a_l\right)z^k\right] \\ &= \sum_{k=n+1}^{\infty} a_k - (1 - z)\sum_{k=n+1}^{\infty}\left(\sum_{l=n+1}^{k} a_l\right)z^k.\end{aligned} \tag{17.38}$$

Now given any $\varepsilon > 0$, there exists an integer $N(\varepsilon) > 0$ such that

$$\left|\sum_{l=n+1}^{k} a_l\right| < \frac{\varepsilon \cos\theta}{6}$$

for all $n > N(\varepsilon)$ and $k > n$, and in particular,

$$\left|\sum_{k=n+1}^{\infty} a_k\right| \leqslant \frac{\varepsilon \cos 0}{6}.$$

Thus it follows from (17.38) that

$$\left|\sum_{k=n+1}^{\infty} a_k(1-z^k)\right| \leqslant \frac{\varepsilon\cos\theta}{6} + \frac{\varepsilon\cos\theta}{6}|1-z|\sum_{k=n+1}^{\infty}|z|^k$$
$$< \frac{\varepsilon\cos\theta}{6} + \frac{\varepsilon}{6}\frac{|1-z|\cos\theta}{1-|z|}. \tag{17.39}$$

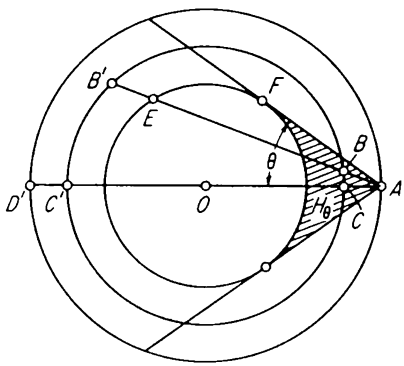

FIGURE 17.7

But near the vertex of G_θ, in the subdomain $H_\theta \subset G_\theta$ bounded by the angle and by the arc of the circle with center at the origin which is tangent to the sides of the angle (see Figure 17.7), we have[15]

$$AB \cdot AB' = AC \cdot AC',$$

and hence

$$\frac{AB}{AC} = \frac{|1-z|}{1-|z|}$$
$$= \frac{AC'}{AB'} < \frac{AD}{AE} < \frac{AD}{AF} = \frac{2}{\cos\theta},$$

i.e.,

$$|1-z|\cos\theta < 2(1-|z|).$$

Therefore

$$\left|\sum_{k=n+1}^{\infty} a_k(1-z^k)\right| < \frac{\varepsilon\cos\theta}{6} + \frac{\varepsilon}{3} < \frac{\varepsilon}{2}$$

for all $n > N(\varepsilon)$ and $z \in H_\theta \subset G_\theta$. Fixing $n > N(\varepsilon)$ and then choosing $\delta(\varepsilon)$ such that

$$\left|\sum_{k=0}^{n} a_k(1-z)^k\right| < \frac{\varepsilon}{2}$$

if $z \in H_\theta$ and $|1-z| < \delta(\varepsilon)$, we deduce from (17.39) that, under the same conditions,

$$|\Delta(z)| = \left|\sum_{k=0}^{\infty} a_k - f(z)\right| < \varepsilon.$$

In other words,

$$\lim_{\substack{z\to 1\\ z\in G_\theta}} f(z) = \sum_{k=0}^{\infty} a_k,$$

as asserted, and the theorem is proved.

---

[15] Here we omit the overbar on line segments, and use the same symbol to denote a line segment and its length. In the figure, B represents the variable point $z \in H_\theta$.

Example 1. Consider the logarithmic series

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{z^n}{n}. \tag{17.40}$$

First we use Theorem 3.6 to prove that this series converges at every point $z = e^{i\theta}$ of the unit circle, except the point $z = -1$. Writing $a_n = (-1)^{n-1} e^{in\theta}$, we observe that

$$|a_1 + \cdots + a_n| = \left| \sum_{k=1}^{n} (-1)^{k-1} e^{ik\theta} \right| = \left| \frac{e^{i\theta} - (-1)^n e^{i(n+1)\theta}}{1 + e^{i\theta}} \right| \leqslant \frac{2}{|1 + e^{i\theta}|},$$

so that the partial sums

$$|a_1 + \cdots + a_n|$$

form a bounded sequence. Then, writing $b_n = 1/n$, we deduce from

$$|b_{n+1} - b_n| = \frac{1}{n} - \frac{1}{n+1}$$

that the series

$$\sum_{n=1}^{\infty} |b_n - b_{n+1}|$$

converges. Since obviously $b_n \to 0$ as $n \to \infty$, it follows from Theorem 3.6 that

$$\sum_{n=1}^{\infty} a_n b_n = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{e^{in\theta}}{n}$$

converges, as asserted.

Moreover, for $|z| < 1$, the sum of the series (17.40) is

$$f(z) = \ln (1 + z) = \ln |1 + z| + i \arg (1 + z)$$

(see Example 1, p. 364). Therefore, according to Theorem 17.14 and Problem 17.23, we have

$$\begin{aligned} &\lim_{z \to e^{i\theta}} [\ln |1 + e^{i\theta}| + i \arg (1 + e^{i\theta})] \\ &\quad = \ln \sqrt{(1 + \cos\theta)^2 + \sin^2\theta} + i \arg \left(e^{i\theta/2} \frac{e^{i\theta/2} + e^{-i\theta/2}}{2} \right) \\ &\quad = \ln \left(2 \cos \frac{\theta}{2} \right) + i \frac{\theta}{2} = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{e^{in\theta}}{n} \end{aligned} \tag{17.41}$$

for any θ in the interval $(-\pi, \pi)$. Separating real and imaginary parts of (17.41), we obtain the following two Fourier expansions, which converge for $-\pi < \theta < \pi$:

$$\ln \left(2 \cos \frac{\theta}{2} \right) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{\cos n\theta}{n},$$

$$\frac{\theta}{2} = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{\sin n\theta}{n}.$$

In particular, the first formula becomes

$$\ln 2 = 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \cdots$$

for $\theta = 0$, and the second formula becomes

$$\frac{\pi}{4} = 1 - \tfrac{1}{3} + \tfrac{1}{5} - \tfrac{1}{7} + \cdots$$

for $\theta = \pi/2$.

Example 2. Consider the binomial series

$$\sum_{n=0}^{\infty} \binom{\alpha}{n} z^n,$$

where we introduce the *binomial coefficient*

$$\binom{\alpha}{n} = \frac{\alpha(\alpha - 1) \cdots (\alpha - n + 1)}{n!}.$$

First suppose $\beta = \operatorname{Re} \alpha > -1$, and write

$$\lambda_n = \left| \binom{\alpha}{n} \Big/ \binom{\alpha}{n-1} \right| = \left| \frac{\alpha - n + 1}{n} \right| = \left| 1 - \frac{1 + \alpha}{n} \right|$$
$$= \left| 1 - \frac{1 + \beta}{n} - \frac{i\gamma}{n} \right| = \sqrt{1 - 2\frac{1 + \beta}{n} + \frac{(1 + \beta)^2 + \gamma^2}{n^2}},$$

where $\gamma = \operatorname{Im} \alpha$. Let β' be a real number satisfying the condition $\beta > \beta' > -1$. Then there is an integer $N(\beta') > 0$ such that for all $n > N(\beta')$,

$$\frac{(1 + \beta^2) + \gamma^2}{n} < 2(\beta - \beta') \quad \text{and} \quad \frac{1 + \beta'}{n} < \frac{1}{2},$$

and hence[16]

$$\lambda_n < \sqrt{1 - 2\frac{1 + \beta}{n} + 2\frac{\beta - \beta'}{n}}$$
$$= \sqrt{1 - 2\frac{1 + \beta'}{n}} < 1 - \frac{1 + \beta'}{n} < \frac{1}{\left(1 + \frac{1}{n}\right)^{1 + \beta'}}$$

or

$$\lambda_n < \frac{n^{1 + \beta'}}{(n + 1)^{1 + \beta'}} = \mu_n. \tag{17.42}$$

---

[16] Since

$$\left(1 + \frac{1}{n}\right)^{-(1 + \beta')} = 1 - \frac{1 + \beta'}{n} + \frac{(1 + \beta')\left(1 + \frac{\beta'}{2}\right)}{n^2} - \frac{(1 + \beta')\left(1 + \frac{\beta'}{2}\right)\left(1 + \frac{\beta'}{3}\right)}{n^3} + \cdots,$$

and since for $(1 + \beta')/n < 1$, the terms of this series decrease in absolute value as n increases, it follows that

$$\left(1 + \frac{1}{n}\right)^{-(1 + \beta')} > 1 - \frac{1 + \beta'}{n}.$$

Letting the index in (17.42) equal $N(\beta') + 1$, $N(\beta') + 2, \ldots, n$, and multiplying the resulting inequalities, we obtain

$$\lambda_{N(\beta')+1}\lambda_{N(\beta')+2} \cdots \lambda_n < \mu_{N(\beta')+1}\mu_{N(\beta')+2} \cdots \mu_n,$$

or

$$\left|\binom{\alpha}{n}\right| < \frac{C(\beta')}{(n+1)^{1+\beta'}}, \tag{17.43}$$

where we have substituted for the λ's and μ's. Here $C(\beta')$ denotes the quantity

$$C(\beta') = \left|\binom{\alpha}{N(\beta')}\right| [N(\beta') + 1]^{1+\beta'},$$

which is independent of n. It follows that the sequence of binomial coefficients

$$\left\{\binom{\alpha}{n}\right\}$$

converges to zero as $n \to \infty$ if $\beta = \operatorname{Re} \alpha > -1$.[17]

Now suppose $\beta = \operatorname{Re} \alpha > 0$, and choose the number β' to be positive, in addition to satisfying the inequality $\beta' < \beta$. Then, since the series

$$\sum_{n=0}^{\infty} \frac{1}{(n+1)^{1+\beta'}}$$

converges for $\beta' > 0$, it follows from (17.43) that the binomial series converges absolutely and uniformly on the unit circle $|z| = 1$. Moreover, for $|z| < 1$, the sum of the binomial series is the function

$$f(z) = (1 + z)^{\alpha} = \exp\{\alpha \ln (1 + z)\}$$

(see Example 2, p. 365). Therefore, according to Theorem 17.14, we have

$$\lim_{z \to e^{i\theta}} (1 + z)^{\alpha} = \sum_{n=0}^{\infty} \binom{\alpha}{n} e^{in\theta}.$$

If $e^{i\theta} = -1$, then

$$\lim_{z \to -1} (1 + z)^{\alpha} = 0.$$

If $e^{i\theta} \neq -1$, i.e., if $-\pi < \theta < \pi$, then

$$\begin{aligned} \lim_{z \to e^{i\theta}} (1 + z)^{\alpha} &= (1 + e^{i\theta})^{\alpha} = \exp\{\alpha \ln [1 + e^{i\theta}]\} \\ &= \exp\left\{\alpha\left[\ln 2 \cos\frac{\theta}{2} + i\frac{\theta}{2}\right]\right\} \\ &= \left(2 \cos\frac{\theta}{2}\right)^{\alpha}\left(\cos\frac{\alpha\theta}{2} + i \sin\frac{\alpha\theta}{2}\right), \end{aligned}$$

and hence

$$\left(2 \cos\frac{\theta}{2}\right)^{\alpha}\left(\cos\frac{\alpha\theta}{2} + i \sin\frac{\alpha\theta}{2}\right) = \sum_{n=0}^{\infty} \binom{\alpha}{n} e^{in\theta}. \tag{17.44}$$

[17] This fact will be used later, in Example 2, p. 408.

In particular, if $\alpha > 0$ is real, taking real and imaginary parts of (17.44), we obtain

$$\left(2\cos\frac{0}{2}\right)^{\alpha}\cos\frac{\alpha 0}{2} = \sum_{n=0}^{\infty}\binom{\alpha}{n}\cos n0,$$

$$\left(2\cos\frac{0}{2}\right)^{\alpha}\sin\frac{\alpha 0}{2} = \sum_{n=0}^{\infty}\binom{\alpha}{n}\sin n0,$$

where, because of the way they are derived, these Fourier expansions converge absolutely and uniformly in the interval $[0, 2\pi]$.

The example of the geometric series

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n$$

shows that the converse of Theorem 17.14 is false. In fact, the limit

$$\lim_{z \to e^{i\theta}} \frac{1}{1-z} = \frac{1}{1-e^{i\theta}}$$

is finite if $e^{i\theta} \neq 1$, whereas

$$\sum_{n=0}^{\infty} e^{in\theta}$$

diverges. However, if suitable restrictions are imposed on the coefficients a_n of the series

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \tag{17.45}$$

(with radius of convergence 1), then the existence of the limit

$$\lim_{r \to 1-} f(re^{i\theta})$$

implies that (17.45) converges for $z = e^{i\theta}$. An important result of this type is given by

THEOREM 17.15 (*Tauber's theorem*). *If the coefficients of the power series* (17.45), *with radius of convergence* 1, *satisfy the condition*

$$\lim_{n \to \infty} na_n = 0, \tag{17.46}$$

and if the limit

$$\lim_{x \to 1-} f(x) = A$$

exists, then the series

$$\sum_{n=0}^{\infty} a_n$$

converges and has sum A.

Proof. According to Problem 3.3, (17.46) also implies

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n} k a_k = 0. \tag{17.47}$$

Consider the quantity

$$\begin{aligned} \sum_{k=0}^{n} a_k - f(x) &= \sum_{k=0}^{n} a_k(1 - x^k) - \sum_{k=n+1}^{\infty} a_k x^k \\ &\leqslant (1 - x) \sum_{k=0}^{n} |a_k|(1 + x + \cdots + x^{k-1}) + \sum_{k=n+1}^{\infty} |a_k| x^k \\ &< n(1 - x) \frac{1}{n} \sum_{k=0}^{n} k|a_k| + \sum_{k=n+1}^{\infty} k|a_k| \frac{x^k}{k}, \end{aligned} \tag{17.48}$$

where $0 < x < 1$. It follows from (17.46) and (17.47) that given any $\varepsilon > 0$, there is an integer $N(\varepsilon) > 0$ such that both

$$n|a_n| < \frac{\varepsilon}{2} \quad \text{and} \quad \frac{1}{n} \sum_{k=0}^{n} k|a_k| < \frac{\varepsilon}{2},$$

provided that $n > N(\varepsilon)$. Then (17.48) implies

$$\begin{aligned} \left| \sum_{k=0}^{n} a_k - f(x) \right| &< n(1 - x) \frac{\varepsilon}{2} + \frac{\varepsilon}{2n} \sum_{k=n+1}^{\infty} x^k \\ &= \frac{\varepsilon}{2} \left[n(1 - x) + \frac{1}{n} \frac{x^{n+1}}{1 - x} \right] \\ &< \frac{\varepsilon}{2} \left[n(1 - x) + \frac{1}{n(1 - x)} \right] \end{aligned}$$

if $n > N(\varepsilon)$. Therefore, setting

$$1 - x = \frac{1}{n},$$

we have

$$\left| \sum_{k=0}^{n} a_k - f\left(1 - \frac{1}{n}\right) \right| < \varepsilon$$

if $n > N(\varepsilon)$, i.e.,

$$\sum_{k=0}^{\infty} a_k = \lim_{n \to \infty} \sum_{k=0}^{n} a_k = \lim_{n \to \infty} f\left(1 - \frac{1}{n}\right) = A,$$

and the theorem is proved.

Remark. The condition (17.46) is unnecessarily strong, and it is not hard to see that it can be replaced by the weaker condition (17.47) [see Problem 17.25]. In fact, Hardy and Littlewood have shown that it is sufficient to assume only that the sequence $\{na_n\}$ is bounded, but the proof of this result

is quite difficult.[18] It can also be shown that replacing (17.46) by the still weaker condition

$$\lim_{n \to \infty} a_n = 0$$

invalidates Tauber's theorem (see, however, Problem 17.26).

According to Theorem 17.11, the circle of convergence of a power series must contain at least one singular point of the sum of the series. This naturally raises the following question: Is there any relation between the distribution of singular points on the circle of convergence of a power series and the structure of the set of points on the circle of convergence where the series converges or diverges? It is easy to see that if there is such a dependence, it cannot be of a simple type. For example, the series

$$f(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^2}$$

has radius of convergence 1, and hence the unit circle contains at least one singular point of $f(z)$, but nevertheless, the series converges absolutely and uniformly on the whole unit circle, and in particular, converges absolutely at the singular point itself. As another example, consider the geometric series

$$\frac{1}{1 - z} = \sum_{n=0}^{\infty} z^n,$$

for which all points of the unit circle other than $z = 1$ are regular points, despite the fact that the series diverges at all points of the unit circle, at the regular points as well as the singular point $z = 1$. Thus we cannot hope to observe any interrelations of the desired type unless certain restrictions are imposed on the class of power series under consideration. The next theorem shows that there is a parallelism between the behavior of a series on its circle of convergence and the distribution of its regular points (provided there are any) if the coefficients of the series approach zero as $n \to \infty$. This requirement already excludes the geometric series, which, as just noted, diverges at its regular points.

THEOREM 17.16 (*Fatou's theorem*). *Let*

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \tag{17.49}$$

be a power series with radius of convergence 1, *such that*

$$\lim_{n \to \infty} a_n = 0.$$

[18] See e.g., E. C. Titchmarsh, *op. cit.*, Sec. 7.66.

Then the series converges (in fact, uniformly) on every arc

$$z = e^{i\theta} \qquad (\alpha \leqslant \theta \leqslant \beta), \tag{17.50}$$

all of whose points are regular points of $f(z)$.[19]

Proof. Let Γ denote the unit circle and K its interior. Suppose $\sigma \subset \Gamma$ is an arc consisting entirely of regular points of $f(z)$. Then for every point $\zeta \in \sigma$ (including the end points of σ) there is a neighborhood $\mathcal{N}(\zeta)$ and an analytic function $\varphi_\zeta(z)$ defined on $\mathcal{N}(\zeta)$ such that $\varphi_\zeta(z) = f(z)$ on $\mathcal{N}(\zeta) \cap K$. Let D be the union of K and all the neighborhoods $\mathcal{N}(\zeta)$, $\zeta \in \sigma$. Then D is a domain containing both K and σ (cf. the proof of Theorem 17.9, where the role of σ is played by the whole circle Γ). If we define a function $\varphi(z)$ on D by setting

$$\varphi(z) = \begin{cases} f(z) & \text{if } z \in K, \\ \varphi_\zeta(z) & \text{if } z \in \mathcal{N}(\zeta), \end{cases}$$

then, just as in the proof of Theorem 17.9, $\varphi(z)$ is single-valued and analytic on D. Since $\varphi(z) = f(z)$ on $K \subset D$, $\varphi(z)$ has the same power series expansion on K as $f(z)$.

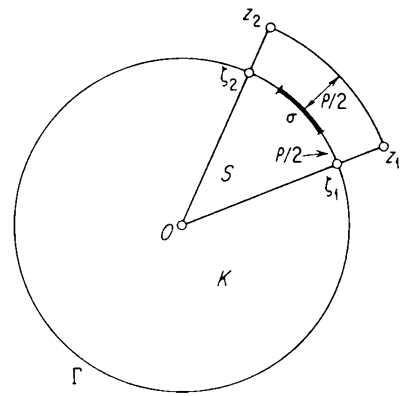

FIGURE 17.8

Now let $\rho > 0$ be the distance between σ and the boundary of D (this boundary contains part of Γ). Moreover, let ζ_1 and ζ_2 be two points which belong to Γ but not to σ, and lie at distance $\rho/2$ from the nearest end point of σ, so that $\sigma \subset \overset{\frown}{\zeta_1\zeta_2} \subset D$. We draw the radii $\overline{O\zeta_1}$ and $\overline{O\zeta_2}$, then extend them a distance $\rho/2$ beyond the circle Γ, and finally join their end points z_1 and z_2 by the arc of a circle of radius $1 + \frac{1}{2}\rho$ (see Figure 17.8). This gives a domain $S = \overset{\frown}{Oz_1z_2}$ (a circular sector), such that $\sigma \subset S$, $\bar{S} \subset D$.

Our problem is to show that the series

$$\sum_{n=0}^{\infty} a_n z^n \tag{17.51}$$

is uniformly convergent on σ. In the course of establishing this fact, it will be shown that

$$\varphi(z) = \sum_{n=0}^{\infty} a_n z^n$$

[19] More precisely, regular points of the element $\{K, f(z)\}$, where K is the unit disk $|z| < 1$. It will be recalled that if there is at least one regular point, then there are whole arcs of the type (17.50) consisting entirely of regular points.

is uniformly convergent on $\bar{S}$. To this end, we consider the function

$$\omega_n(z) = \frac{\varphi_n(z)}{z^{n+1}}(z - \zeta_1)(z - \zeta_2),$$

where

$$\varphi_n(z) = \varphi(z) - \sum_{k=0}^{n} a_k z^k.$$

Since

$$\omega_n(z) = \frac{1}{z^{n+1}}\left(\sum_{k=0}^{\infty} a_k z^k - \sum_{k=0}^{n} a_k z^k\right)(z - \zeta_1)(z - \zeta_2) \qquad (17.52)$$

$$= \left(\sum_{k=n+1}^{\infty} a_k z^{k-n-1}\right)(z - \zeta_1)(z - \zeta_2) \qquad (z \in K),$$

it is clear that $\omega_n(z)$ is analytic on K, and indeed on D. We intend to estimate the function $|\omega_n(z)|$ on the boundary B of the sector S, and then use the maximum modulus principle to estimate $|\omega_n(z)|$ on S itself. The advantage of the function $\omega_n(z)$ over the function $\varphi_n(z)$ is that for sufficiently large n, the denominator z^{n+1} makes $|\omega_n(z)|$ small if $z \in B \cap E(\Gamma)$, i.e., at points of B belonging to the exterior of Γ, whereas the factor $(z - \zeta_1)(z - \zeta_2)$ makes $|\omega_n(z)|$ small at points of the radii $\overline{Oz_1}$ and $\overline{Oz_2}$ near ζ_1 or ζ_2. We begin by estimating $|\omega_n(z)|$ on the set consisting of the segment $\overline{O\zeta_1}$ (or $\overline{O\zeta_2}$) without its end points. Let z belong to this set (which is a subset of K), let $\varepsilon > 0$ be arbitrary, and let $N(\varepsilon) > 0$ be an integer such that $|a_n| < \varepsilon$ for all $n > N(\varepsilon)$. Then

$$|\omega_n(z)| = \left|\left(\sum_{k=n+1}^{\infty} a_k z^{k-n-1}\right)(z - \zeta_1)(z - \zeta_2)\right|$$

$$\leqslant \varepsilon \sum_{k=n+1}^{\infty} |z|^{k-n-1}|z - \zeta_1|\,|z - \zeta_2| < 2\varepsilon$$

if $n > N(\varepsilon)$ [see (17.52)], where we use the obvious relations

$$\sum_{k=n+1}^{\infty} |z|^{k-n-1} = \frac{1}{1 - |z|}, \quad |z - \zeta_1| = 1 - |z|,$$

$$|z - \zeta_2| \leqslant |z| + |\zeta_2| < 2.$$

Next let z belong to the set consisting of the segment $\overline{\zeta_1 z_1}$ (or $\overline{\zeta_2 z_2}$) without its end points. Then, noting that

$$|z - \zeta_1| = |z| - 1, \qquad |z - \zeta_2| \leqslant |z| + |\zeta_2| < 2R,$$

where $R = 1 + \frac{1}{2}\rho$, and writing

$$M = \max_{z \in \bar{S}} |\varphi(z)|,$$

we have

$$
\begin{aligned}
|\omega_n(z)| &= \left|\frac{1}{z^{n+1}}\left(\varphi(z) - \sum_{k=0}^{n} a_k z^k\right)(z-\zeta_1)(z-\zeta_2)\right| \\
&\leqslant \frac{1}{|z|^{n+1}}\left(M + \sum_{k=0}^{n} |a_k|\,|z|^k\right)(|z|-1)2R \\
&= 2R\left(M + \sum_{k=0}^{N(\varepsilon)} |a_k|\,|z|^k\right)\frac{|z|-1}{|z|^{n+1}} + 2R\left(\sum_{k=N(\varepsilon)+1}^{n} |a_k|\,|z|^k\right)\frac{|z|-1}{|z|^{n+1}} \\
&< 2R\left(M + \sum_{k=0}^{N(\varepsilon)} |a_k|R^k\right)\frac{|z|-1}{|z|^{n+1}} + 2R\varepsilon \sum_{k=N(\varepsilon)+1}^{n} |z|^k \frac{|z|-1}{|z|^{n+1}} \\
&= 2R\left(M + \sum_{k=0}^{N(\varepsilon)} |a_k|R^k\right)\frac{|z|-1}{|z|^{n+1}} + 2R\varepsilon \frac{|z|^{n+1} - |z|^{N(\varepsilon)+1}}{|z|-1}\frac{|z|-1}{|z|^{n+1}} \\
&< 2R\left(M + \sum_{k=0}^{N(\varepsilon)} |a_k|R^k\right)\frac{|z|-1}{|z|^{n+1}} + 2R\varepsilon,
\end{aligned}
$$

provided that $n > N(\varepsilon)$. But

$$
\frac{|z|-1}{|z|^{n+1}} < \frac{|z|-1}{|z|^{n+1}-1} = \frac{1}{|z|^n + \cdots + |z| + 1} < \frac{1}{n+1},
$$

and hence

$$
|\omega_n(z)| < \frac{2R}{n+1}\left(M + \sum_{k=0}^{N(\varepsilon)} |a_k|R^k\right) + 2R\varepsilon.
$$

Choosing an integer $N_1(\varepsilon) \geqslant N(\varepsilon)$ such that the first term in the right-hand side of this inequality is less than ε if $n > N_1(\varepsilon)$, we have

$$
|\omega_n(z)| < (2R+1)\varepsilon \qquad \text{if} \quad n > N_1(\varepsilon).
$$

Now let z belong to the arc $\overset{\frown}{z_1 z_2}$ (including its end points). In this case,

$$
|z-\zeta_1| < 2R, \qquad |z-\zeta_2| < 2R,
$$

and

$$
\begin{aligned}
|\omega_n(z)| &< \frac{1}{R^{n+1}}\left(M + \sum_{k=0}^{n} |a_k|R^k\right)4R^2 \\
&= \frac{4}{R^{n-1}}\left(M + \sum_{k=0}^{N(\varepsilon)} |a_k|R^k\right) + \frac{4}{R^{n-1}} \sum_{k=N(\varepsilon)+1}^{n} |a_k|R^k \\
&\leqslant \frac{4}{R^{n-1}}\left(M + \sum_{k=0}^{N(\varepsilon)} |a_k|R^k\right) + \frac{4\varepsilon}{R^{n-1}} \sum_{k=N(\varepsilon)+1}^{n} R^k \\
&< \frac{4}{R^{n-1}}\left(M + \sum_{k=0}^{N(\varepsilon)} |a_k|R^k\right) + \frac{4R^2\varepsilon}{R-1} \\
&= \frac{4}{R^{n-1}}\left(M + \sum_{k=0}^{N(\varepsilon)} |a_k|R^k\right) + \frac{8R^2\varepsilon}{\rho}.
\end{aligned}
$$

Choosing an integer $N_2(\varepsilon) \geqslant N_1(\varepsilon) \geqslant N(\varepsilon)$ such that the first term in the right-hand side of this inequality is less than ε if $n > N_2(\varepsilon)$, we have

$$|\omega_n(z)| < \left(\frac{8R^2}{\rho} + 1\right)\varepsilon \qquad \text{if} \quad n > N_2(\varepsilon).$$

At the point $z = 0$, $\omega_n(z)$ takes the value $a_{n+1}\zeta_1\zeta_2$ [equal to $\lim_{z \to 0} \omega_n(z)$], and hence

$$|\omega_n(0)| = |a_{n+1}| < \varepsilon \qquad \text{if} \quad n > N(\varepsilon).$$

Moreover, $\omega_n(z)$ vanishes at the points $z = \zeta_1$ and $z = \zeta_2$. Finally, therefore, if $n > N_2(\varepsilon)$,

$$|\omega_n(z)| < \max\left\{2\varepsilon, (2R + 1)\varepsilon, \left(\frac{8R^2}{\rho} + 1\right)\varepsilon\right\} = \left(\frac{8R^2}{\rho} + 1\right)\varepsilon$$

at all points of the boundary B of the sector S. It follows from the maximum modulus principle (Theorem 17.5) that if $n > N_2(\varepsilon)$, then

$$|\omega_n(z)| < \left(\frac{8R^2}{\rho} + 1\right)\varepsilon$$

at all points of S, in particular at all points of the arc σ. But on σ we have

$$|\omega_n(z)| = \left|\varphi(z) - \sum_{k=0}^{n} a_k z^k\right| |z - \zeta_1|\,|z - \zeta_2| > \left|\varphi(z) - \sum_{k=0}^{n} a_k z^k\right| \left(\frac{\rho}{2}\right)^2,$$

and hence

$$\left|\varphi(z) - \sum_{k=0}^{n} a_k z^k\right| < \frac{4(8R^2 + \rho)\varepsilon}{\rho^3} \qquad \text{if} \quad n > N_2(\varepsilon).$$

Since $\varepsilon > 0$ is arbitrarily small, the theorem is proved.

Example 1. Consider the logarithmic series

$$\ln(1 + z) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{z^n}{n},$$

and let $\sigma: z = e^{i\theta}$ $(\alpha \leqslant 0 \leqslant \beta)$ be any arc which does not contain the point $z = -1$, the only singular point of $\ln(1 + z)$ on the unit circle. Since

$$\lim_{n \to \infty} \frac{(-1)^{n-1}}{n} = 0,$$

it follows from Fatou's theorem that the logarithmic series is uniformly convergent on σ.

Example 2. Consider the binomial series

$$(1 + z)^{\alpha} = \sum_{n=0}^{\infty} \binom{\alpha}{n} z^n,$$

and let σ be the same as in Example 1.[20] As shown in Example 2, p. 400,

$$\lim_{n \to \infty} \binom{\alpha}{n} = 0$$

if $\operatorname{Re} \alpha > -1$. Therefore, according to Fatou's theorem, the binomial series is uniformly convergent on σ if $\operatorname{Re} \alpha > -1$.

Example 3. The fact that the coefficients of a power series with radius of convergence 1 converge to zero, while the series converges uniformly on some arc σ of the unit circle Γ, does not imply that the points of σ are regular. This is shown by the series

$$\sum_{n=0}^{\infty} \frac{z^{2^n}}{2^{n^2}},$$

for which every point of Γ is a singular point, even though the series is uniformly convergent on Γ (see the example on p. 390).

Example 4. Finally, following Luzin, we construct a power series with radius of convergence 1 which diverges at every point of the unit circle Γ, despite the fact that its coefficients converge to zero. Consider the polynomial

$$g_p(z) = 1 + z + \cdots + z^{p-1} = \frac{1 - z^p}{1 - z},$$

where $p > 0$ is an integer. At the point $z = e^{i\theta}$ $(\theta \neq 0)$ of Γ we have

$$|g_p(e^{i\theta})| = \left|\frac{1 - e^{ip\theta}}{1 - e^{i\theta}}\right| = \left|\frac{e^{ip\theta/2}(e^{ip\theta/2} - e^{-ip\theta/2})}{e^{i\theta/2}(e^{i\theta/2} - e^{-i\theta/2})}\right| = \left|\frac{\sin(p\theta/2)}{\sin(\theta/2)}\right|.$$

Since

$$\frac{2}{\pi}|\varphi| \leqslant |\sin \varphi| \leqslant |\varphi|$$

if $|\varphi| \leqslant \pi/2$ [cf. (13.20)], it follows that

$$|g_p(e^{i\theta})| \geqslant \frac{(2/\pi)(p|\theta|/2)}{|\theta|/2} = \frac{2}{\pi} p \tag{17.53}$$

if $0 < |\theta| \leqslant \pi/p$. Moreover, since $g_p(1) = p$, (17.53) also holds for $\theta = 0$. It is easy to see that given an arbitrary point $\zeta_0 = e^{i\theta}$ of Γ, there exists an integer m_0, $0 \leqslant m_0 \leqslant p - 1$ such that the point $z = \zeta_0 e^{-2\pi i m_0/p}$ belongs to an arc of Γ on which $|\arg z| \leqslant \pi/p$. Then

$$|g_p(\zeta_0 e^{-2\pi i m_0/p})| \geqslant \frac{2}{\pi} p,$$

and hence

$$\max_{0 \leqslant m \leqslant p-1} |g_p(\zeta_0 e^{-2\pi i m/p})| \geqslant \frac{2}{\pi} p. \tag{17.54}$$

---

[20] It is assumed that α is not a nonnegative integer (otherwise the binomial series reduces to a polynomial). It is easy to see that then $z = -1$ is the only singular point of $(1 + z)^\alpha$ on the unit circle.

Next, for every integer $p > 0$, we form the polynomial

$$H_p(z) = g_p(z) + z^p g_p(ze^{-2\pi i/p}) + z^{2p} g_p(ze^{-2\pi i(2/p)}) + \cdots + z^{(p-1)p} g_p(ze^{-2\pi i(p-1)/p}).$$

It is clear that $g_p(z)$ contains all powers of z from 0 to $p-1$, $z^p g_p(ze^{-2\pi i/p})$ contains all powers of z from p to $2p-1$, and so on, up to $z^{(p-1)p} g_p(ze^{-2\pi i(p-1)/p})$, which contains all powers of z from $(p-1)p$ to p^2-1. Therefore $H_p(z)$ is a polynomial of degree p^2-1 in z, all of whose coefficients have absolute value 1. Using the polynomials $H_p(z)$, we form *Luzin's series*

$$H_1(z) + \sum_{p=2}^{\infty} \frac{1}{\sqrt{p}} z^{1^2+2^2+\cdots+(p-1)^2} H_p(z). \tag{17.55}$$

Since the pth term of (17.55) is a polynomial containing powers of z from $1^2 + 2^2 + \cdots + (p-1)^2$ to $1^2 + 2^2 + \cdots + (p-1)^2 + p^2 - 1$, no two terms of the series contain the same powers of z. Thus, if we write (17.55) in the form

$$\sum_{n=0}^{\infty} a_n z^n, \tag{17.56}$$

it is an easy matter to determine the coefficients a_n $(n = 0, 1, 2, \ldots)$. In particular,

$$|a_n| = \frac{1}{\sqrt{p}}$$

for all n from $1^2 + 2^2 + \cdots + (p-1)^2$ to $1^2 + 2^2 + \cdots + (p-1)^2 + p^2 - 1$. Therefore

$$\lim_{n\to\infty} a_n = 0, \tag{17.57}$$

and moreover, the series (17.56) converges for $|z| < 1$, by the ratio test. However, the series diverges at every point of $\Gamma: |z| = 1$, since the series

$$\sum_{n=0}^{\infty} a_n z^n = \sum_{p=1}^{\infty} \left[\sum_{m=0}^{p-1} \frac{1}{\sqrt{p}} z^{1^2+2^2+\cdots+(p-1)^2} z^{mp} g_p(ze^{-2\pi im/p}) \right]$$

does not satisfy the Cauchy convergence criterion for any $z = \zeta \in \Gamma$. In fact, according to (17.54),

$$\max_{0\le m\le p-1} \left| \frac{1}{\sqrt{p}} \zeta^{1^2+2^2+\cdots+(p-1)^2} \zeta^{mp} g_p(\zeta e^{-2\pi im/p}) \right| = \max_{0\le m\le p-1} \frac{1}{\sqrt{p}} \left| g_p(\zeta e^{-2\pi im/p}) \right| \frac{2p}{\pi\sqrt{p}} \ge \frac{2}{\pi}\sqrt{p},$$

i.e., blocks of consecutive terms whose sums have arbitrarily large absolute values appear infinitely often in Luzin's series. Furthermore, it follows from Fatou's theorem that every point of Γ is a singular point of Luzin's series, since, because of (17.57), the series must converge at regular points.

In conclusion, we note that the series

$$a_0 - a_0 z + a_1 z^2 - a_1 z^3 + \cdots + a_n z^{2n} - a_n z^{2n+1} + \cdots, \qquad (17.58)$$

with the same coefficients a_n as Luzin's series, converges at the point $z = 1$ [because of its "telescoping structure" and (17.57)], but diverges at every other point of the unit circle Γ. In fact, if we assume that (17.58) converges at a point $\zeta \neq 1$, $\zeta \in \Gamma$, then the series

$$a_0(1 - \zeta) + a_1\zeta^2(1 - \zeta) + \cdots + a_n\zeta^{2n}(1 - \zeta) + \cdots$$

must also converge. But this implies that the series

$$a_0 + a_1\zeta^2 + \cdots + a_n\zeta^{2n} + \cdots$$

converges, i.e., that Luzin's series converges at the point $\zeta^2 \in \Gamma$, which is impossible.

*86. Compact Families of Analytic Functions

A point set is said to be *conditionally compact* if every sequence $\{z_n\}$ of points $z_n \in E$ contains a convergent subsequence $\{z_{n_k}\}$. As applied to point sets, the concept of conditional compactness is superfluous, since the class of conditionally compact sets is precisely the class of *bounded* sets. In fact, if E is bounded, then E is conditionally compact, by Theorem 3.3, whereas if E is conditionally compact, then E must be bounded, since otherwise, for every positive integer n there is a point $z_n \in E$ such that $|z_n| > n$, which means that the sequence $\{z_n\}$ cannot have a convergent subsequence. A set which is both closed and conditionally compact is said to be *compact*. This explains the presence of the adverb "conditionally," and is consistent with our earlier definition of a compact point set as a closed bounded set (cf. Problem 3.23).

The following definition ascribes a precise technical meaning to the preposition "inside," as applied to a domain, and allows us to keep our language simple:

DEFINITION. *A property which holds on every compact subset of a domain G is said to hold inside G.*

When we turn to sets whose elements are *functions*, the concept of conditional compactness acquires more importance. Let $\mathcal{F}$ be a set (in general, infinite) of functions, all of which are analytic on the same domain G. Such a set is called a *family of analytic functions on G* (this definition includes the case where $\mathcal{F}$ is a sequence of functions, all analytic on G). Given a domain G and a family $\mathcal{F}$ of analytic functions on G, we say that $\mathcal{F}$ is *conditionally compact* (*on G*) if every sequence $\{f_n(z)\}$ of functions $f_n(z) \in \mathcal{F}$ contains a subsequence $\{f_{n_k}(z)\}$ which is uniformly convergent *inside* G. In

this context, the word "conditionally" is usually omitted for simplicity of language, but the distinction between this use of the word "compact" and the previous use must be kept clearly in mind. The importance of the concept of a compact family of analytic functions on G stems from the fact that the limit function

$$f(z) = \lim_{k \to \infty} f_{n_k}(z)$$

is itself analytic on G (see Theorem 15.8), a property which is often very useful.

Example 1. The sequence of functions

$$z, 2z, \ldots, nz, \ldots,$$

all analytic in the whole finite plane, is not compact on any domain G.

Example 2. The sequence of functions

$$z, z^2, \ldots, z^n, \ldots,$$

all analytic in the whole finite plane, is compact on the disk $|z| < 1$, but not on the disk $|z| < 2$. In fact, $z^n \to 0$ uniformly for $|z| < 1$, but $z^n \to \infty$ for $|z| > 1$.

The basic result to be proved in this section is a necessary and sufficient condition for compactness of families of analytic functions. To establish this result, we shall need the following

LEMMA. *Let* $\{f_n(z)\}$ *be a sequence of functions, all of which are analytic on the disk* $K: |z - z_0| < R$. *Suppose* $\{f_n(z)\}$ *is uniformly bounded on* K, *i.e., suppose there exists a constant* $M > 0$ *such that*

$$|f_n(z)| \leqslant M$$

for all n *and* $z \in K$ (*cf. p.* 331). *Then* $\{f_n(z)\}$ *is compact on* K.

Proof. Every function $f_n(z)$ has a Taylor series expansion

$$f_n(z) = a_0^{(n)} + a_1^{(n)}(z - z_0) + \cdots + a_k^{(n)}(z - z_0)^k + \cdots \quad (17.59)$$

on K. According to the Cauchy inequalities [cf. (16.43)], we have

$$|a_k^{(n)}| \leqslant \frac{M_n(\rho)}{\rho^k}$$

for every ρ $(0 < \rho < R)$, where

$$M_n(\rho) = \max_{|z - z_0| = \rho} |f_n(z)| \leqslant M.$$

Letting ρ approach the limit R, we find that

$$|a_k^{(n)}| \leqslant \frac{M}{R^k} \quad (17.60)$$

for all $n > 0$ and $k \geqslant 0$. It follows that for every fixed k, the sequence $\{a_k^{(n)}\}$ is bounded. Using this fact, from the sequence $\{f_n(z)\}$, we choose a subsequence

$$f_{n_1^{(1)}}(z), f_{n_2^{(1)}}(z), \ldots, f_{n_m^{(1)}}(z), \ldots, \tag{17.61}$$

such that the sequence $\{a_0^{(n_m^{(1)})}\}$, consisting of the coefficients of the zeroth powers of $z - z_0$ in the Taylor series of the functions (17.61), converges to a finite limit as $m \to \infty$. Then from the subsequence (17.61) we choose a new subsequence

$$f_{n_1^{(2)}}(z), f_{n_2^{(2)}}(z), \ldots, f_{n_m^{(2)}}(z), \ldots, \tag{17.62}$$

such that the sequence $\{a_1^{n_m^{(2)}}\}$, consisting of the first powers of $z - z_0$ in the Taylor series of the functions (17.62), converges as $m \to \infty$. Continuing in this way, suppose we have chosen a subsequence

$$f_{n_1^{(N)}}(z), f_{n_2^{(N)}}(z), \ldots, f_{n_m^{(N)}}(z), \ldots, \tag{17.63}$$

such that all N sequences

$$\{a_0^{n_m^{(N)}}\}, \{a_1^{n_m^{(N)}}\}, \ldots, \{a_{N-1}^{n_m^{(N)}}\}$$

converge as $m \to \infty$. Then from (17.63) we can choose a new subsequence

$$f_{n_1^{(N+1)}}(z), f_{n_2^{(N+1)}}(z), \ldots, f_{n_m^{(N+1)}}(z), \ldots,$$

such that all $N + 1$ sequences

$$\{a_0^{n_m^{(N+1)}}\}, \{a_1^{n_m^{(N+1)}}\}, \ldots, \{a_N^{n_m^{(N+1)}}\}, \ldots$$

converge as $m \to \infty$. Thus our construction leads to the following infinite family of subsequences

$$f_{n_1^{(l)}}(z), f_{n_2^{(l)}}(z), \ldots, f_{n_m^{(l)}}(z), \ldots \quad (l = 1, 2, \ldots). \tag{17.64}$$

We now form a new sequence, called the *diagonal sequence* of the family (17.64), whose lth term ($l = 1, 2, \ldots$) is the term standing in the lth position of the lth subsequence. Thus the diagonal sequence is

$$f_{n_1^{(1)}}(z), f_{n_2^{(2)}}(z), \ldots, f_{n_m^{(m)}}(z), \ldots,$$

which we abbreviate to

$$f_{\nu_1}(z), f_{\nu_2}(z), \ldots, f_{\nu_m}(z), \ldots, \tag{17.65}$$

by writing $\nu_m = n_m^{(m)}$ ($m = 1, 2, \ldots$). Obviously, (17.65) is a subsequence of the original sequence $\{f_n(z)\}$, and

$$f_{\nu_m}(z) = a_0^{(\nu_m)} + a_1^{(\nu_m)}(z - z_0) + \cdots + a_k^{(\nu_m)}(z - z_0)^k + \cdots.$$

Moreover, given any nonnegative integer k, all the functions in (17.65), starting from the $(k+1)$th function $f_{\nu_{k+1}}(z)$, belong to the subsequence

$$f_{n_1^{(k+1)}}(z), f_{n_2^{(k+1)}}(z), \ldots, f_{n_m^{(k+1)}}(z), \ldots,$$

and hence the sequence $\{a_k^{(\nu_m)}\}$ converges. If we set

$$\lim_{m\to\infty} a_k^{(\nu_m)} = a_k,$$

then, since each coefficient $a_k^{(\nu_m)}$ satisfies the inequality (17.60), the same is true of the coefficients a_k:

$$|a_k| \leqslant \frac{M}{R^k} \qquad (k = 0, 1, 2, \ldots). \tag{17.66}$$

It follows that

$$\overline{\lim_{k\to\infty}} \sqrt[k]{|a_k|} \leqslant \frac{1}{R},$$

and therefore, by Theorem 16.1, the radius of convergence of the series

$$a_0 + a_1(z - z_0) + \cdots + a_k(z - z_0)^k + \cdots \tag{17.67}$$

is no less than R. This means that the sum $f(z)$ of the series (17.67) is an analytic function on the disk K.

We now show that the sequence (17.65) converges uniformly to $f(z)$ inside K. As usual (see p. 348), it is sufficient to show that (17.65) converges uniformly to $f(z)$ on every closed disk $|z - z_0| \leqslant \rho < R$. Given any $\varepsilon > 0$, there is an integer $n_0 > 0$ such that

$$\sum_{n=n_0+1}^{\infty} M\left(\frac{\rho}{R}\right)^n < \frac{\varepsilon}{3}.$$

Then, because of the inequalities (17.60) and (17.66), we have

$$\sum_{n=n_0+1}^{\infty} |a_n^{(\nu_m)}|\rho^n \leqslant \sum_{n=n_0+1}^{\infty} M\left(\frac{\rho}{R}\right)^n < \frac{\varepsilon}{3},$$

$$\sum_{n=n_0+1}^{\infty} |a_n|\rho^n \leqslant \sum_{n=n_0+1}^{\infty} M\left(\frac{\rho}{R}\right)^n < \frac{\varepsilon}{3},$$

and hence, for $|z - z_0| \leqslant \rho$,

$$\begin{aligned} |f_{\nu_m}(z) - f(z)| &= \left|\sum_{n=0}^{\infty} (a_n^{(\nu_m)} - a_n)(z - z_0)^n\right| \\ &\leqslant \sum_{n=0}^{n_0} |a_n^{(\nu_m)} - a_n|\rho^n + \sum_{n=n_0+1}^{\infty} |a_n^{(\nu_m)}|\rho^n + \sum_{n=n_0+1}^{\infty} |a_n|\rho^n \\ &< \sum_{n=0}^{n_0} |a_n^{(\nu_m)} - a_n|\rho^n + \frac{2\varepsilon}{3}. \end{aligned} \tag{17.68}$$

Since n_0 is fixed and

$$\lim_{m\to\infty} a_n^{(\nu_m)} = a_n,$$

it follows that

$$\sum_{n=0}^{n_0} |a_n^{(\nu_m)} - a_n|\rho^n < \frac{\varepsilon}{3}$$

for sufficiently large $m > m_0$. But then, according to (17.68),

$$|f_{\nu_m}(z) - f(z)| < \varepsilon$$

if $|z - z_0| \leqslant \rho$ and $m > m_0$. Since ρ $(0 < \rho < R)$ is arbitrary, the sequence (17.65) converges uniformly to $f(z)$ inside K, and the lemma is proved.

DEFINITION. *Let $\mathscr{E}$ be a set of functions, all defined on the same domain G, and suppose that for every closed bounded subset $F \subset G$,[21] there is a number $M(F) > 0$ such that*

$$|f(z)| \leqslant M(F)$$

for all $f(z) \in \mathscr{E}$ and $z \in F$. Then $\mathscr{E}$ is said to be uniformly bounded inside G.

THEOREM 17.17 (*Montel's compactness criterion*). *Let G be a domain, and $\mathscr{E}$ a family of analytic functions on G. Then a necessary and sufficient condition for $\mathscr{E}$ to be compact is that $\mathscr{E}$ be uniformly bounded inside G.*

Proof. To show that the condition is necessary, suppose $\mathscr{E}$ is compact but not uniformly bounded inside G. Then there exists a closed bounded subset $F \subset G$ on which suitable functions in $\mathscr{E}$ have arbitrarily large absolute values. In other words, for every positive integer n, we can find a function $f_n(z) \in \mathscr{E}$ and a point $z_n \in F$ such that

$$|f_n(z_n)| > n \qquad (n = 1, 2, \ldots).$$

On the other hand, since $\mathscr{E}$ is compact, $\{f_n(z)\}$ contains a subsequence $\{f_{n_k}(z)\}$ which is uniformly convergent inside G, in particular, uniformly convergent on F. The limit function

$$f(z) = \lim_{k\to\infty} f_{n_k}(z)$$

is analytic on G, and hence continuous on F, which implies that

$$\max_{z\in F} |f(z)| = M < \infty.$$

---

[21] Here we avoid using the word *compact* as a synonym for *closed and bounded*, since we shall also be talking about (conditionally) compact families of analytic functions.

Since $\{f_{n_k}(z)\}$ converges uniformly to $f(z)$ on F, we must have

$$|f_{n_k}(z) - f(z)| < 1$$

or

$$|f_{n_k}(z)| < |f(z)| + 1 \leqslant M + 1$$

for all $z \in F$ and sufficiently large $n_k > N$. But this contradicts the fact that the numbers $|f_{n_k}(z_{n_k})|$ become arbitrarily large (in fact, exceed n_k), and therefore the condition is necessary.

To show that the condition is sufficient, suppose $\mathscr{F}$ is uniformly bounded inside G, and let $\{f_n(z)\}$ be any sequence of functions $f_n(z) \in \mathscr{F}$. First we prove that given any closed bounded subset $F \subset G$, $\{f_n(z)\}$ contains a subsequence which converges uniformly on F.[22] To show this, given any $z \in F$, let the neighborhood $\mathscr{N}(z)$ be such that $\overline{\mathscr{N}(z)} \subset G$, and let $\mathscr{N}^*(z)$ be any "smaller" neighborhood of z, i.e., let $\mathscr{N}^*(z)$ be strictly included in $\mathscr{N}(z)$. By the Heine-Borel theorem (Theorem 3.14), there exists a finite number of points $z_1, \ldots, z_q$ in F such that

$$F \subset \mathscr{N}^*(z_1) \cup \cdots \cup \mathscr{N}^*(z_q). \tag{17.69}$$

By hypothesis, the sequence $\{f_n(z)\}$ is uniformly bounded inside G, and hence on every neighborhood $\mathscr{N}(z_j)$, $j = 1, \ldots, q$, which means that

$$|f_n(z)| \leqslant M_j \quad \text{if} \quad z \in \mathscr{N}(z_j) \qquad (n = 1, 2, \ldots).$$

Therefore, according to the lemma proved above, $\{f_n(z)\}$ contains a subsequence $\{f_{r_m^{(1)}}(z)\}$ which converges uniformly inside $\mathscr{N}(z_1)$, $\{f_{r_m^{(1)}}(z)\}$ contains a subsequence $\{f_{r_m^{(2)}}(z)\}$ which converges uniformly inside $\mathscr{N}(z_2)$, and so on, until finally, $\{f_{r_m^{(q-1)}}(z)\}$ contains a subsequence $\{f_{r_m^{(q)}}(z)\}$ which converges uniformly inside $\mathscr{N}(z_q)$. Moreover, because of the way it is constructed, the last of the subsequences, i.e., $\{f_{r_m^{(q)}}(z)\}$, converges uniformly *inside* every neighborhood $\mathscr{N}(z_j)$, $j = 1, \ldots, q$, and in particular, converges uniformly *on* every neighborhood $\mathscr{N}^*(z_j)$, $j = 1, \ldots, q$. Therefore, because of (17.69), $\{f_{r_m^{(q)}}(z)\}$ converges uniformly on the set F.

The final step is to show that $\{f_n(z)\}$ contains a subsequence which converges uniformly inside G, i.e., on every closed bounded subset $F \subset G$. With this in mind, let $\{F_n\}$ be exactly the same increasing sequence of closed bounded subsets of G approximating G constructed in the proof of Theorem 17.3. Clearly, every closed bounded $F \subset G$ is contained in all the F_n starting from a certain value of n ($F_n \subset F_{n+1}$ for all $n = 1, 2, \ldots$). In fact, let Γ be the boundary of G, and let $\rho > 0$ be the

[22] Later, we shall prove that $\{f_n(z)\}$ contains a subsequence which converges uniformly on *every* closed bounded subset $F \subset G$.

distance between F and Γ; then $F \subset F_n$ if $n > 1/\rho$. From the sequence $\{f_n(z)\}$ we choose a subsequence $\{f_{s_m^{(1)}}(z)\}$ which converges uniformly on F_1 (the existence of such a subsequence was just proved), from $\{f_{s_m^{(1)}}(z)\}$ we choose a subsequence $\{f_{s_m^{(2)}}(z)\}$ which converges uniformly on F_2, and so on. In this way, we obtain an infinite family of subsequences

$$\{f_{s_m^{(l)}}(z)\} \qquad (l = 1, 2, \ldots), \tag{17.70}$$

each contained in all the subsequences with smaller values of l, where the lth subsequence converges uniformly on F_l. The diagonal sequence corresponding to the family (17.70), i.e., the sequence

$$f_{s_1^{(1)}}(z), f_{s_2^{(2)}}(z), \ldots, f_{s_m^{(m)}}(z), \ldots, \tag{17.71}$$

is obviously a subsequence of the original sequence $\{f_n(z)\}$, and moreover, given any positive integer k, all the functions in (17.71), starting from the kth function $f_{s_k^{(k)}}(z)$, belong to the subsequence $\{f_{s_m^{(k)}}(z)\}$. Therefore (17.71) converges uniformly on every set F_n, and consequently on every closed bounded $F \subset G$ (since, as already noted, every such F is contained in all the F_n starting from a certain value of n). In other words, every sequence $\{f_n(z)\}$ of functions $f_n(z) \in \mathcal{F}$ contains a subsequence which converges uniformly inside G, i.e., the family $\mathcal{F}$ is compact. This completes the proof of Montel's compactness criterion.

*87. Vitali's Theorem. Analytic Functions Defined by Integrals

The following theorem establishes an important property of compact families of analytic functions:

THEOREM 17.18 (*Vitali's theorem*). *Let G be a domain, and let $\{f_n(z)\}$ be a compact sequence of analytic functions on G. Suppose $\{f_n(z)\}$ converges on a set of points $E \subset G$ with a limit point in G. Then $\{f_n(z)\}$ converges uniformly inside G.*

Proof. Since the sequence $\{f_n(z)\}$ is compact, any of its subsequences $\{f_{n_k}(z)\}$ contains another subsequence $\{f_{n_k^*}(z)\}$ which converges uniformly inside G. First we show that all the subsequences of $\{f_n(z)\}$ which converge uniformly inside G converge to the same limit function $f(z)$. In fact, let $\{f_{\mu_k}(z)\}$ and $\{f_{\nu_k}(z)\}$ be two such subsequences of $\{f_n(z)\}$, and suppose that

$$\lim_{k \to \infty} f_{\mu_k}(z) = f(z), \qquad \lim_{k \to \infty} f_{\nu_k}(z) = \varphi(z).$$

By Theorem 15.8, the functions $f(z)$ and $\varphi(z)$ are analytic on G. Moreover, since $\{f_n(z)\}$ converges on E to the function

$$\lim_{n \to \infty} f_n(z), \tag{17.72}$$

the subsequences $\{f_{\mu_k}(z)\}$ and $\{f_{\nu_k}(z)\}$ must also converge to (17.72) on E, and hence must coincide on E. Therefore, since E has a limit point in G, it follows from the interior uniqueness theorem (Theorem 17.1) that $f(z) = \varphi(z)$ on the whole domain G.

Next we show that the original sequence $\{f_n(z)\}$ converges uniformly inside G to the function $f(z)$ just constructed. Suppose this is not the case. Then we can find a closed bounded subset $F \subset G$, a sequence $\{n_k\}$ of positive integers (where $n_1 < \cdots < n_k < \cdots$), and a sequence $\{z_k\}$ of points in F such that

$$|f_{n_k}(z_k) - f(z_k)| \geqslant \alpha > 0 \qquad (k = 1, 2, \ldots). \tag{17.73}$$

Consider the sequence $\{f_{n_k}(z)\}$. It must contain a subsequence $\{f_{n'_k}(z)\}$ which converges uniformly on F to the limit function $f(z)$. Therefore

$$|f_{n'_k}(z) - f(z)| < \alpha$$

for all $z \in F$ and sufficiently large $n'_k > N$. In particular, we have

$$|f_{n'_k}(z_k) - f(z_k)| < \alpha \tag{17.74}$$

for all $n'_k > N$. But (17.74) contradicts (17.73), and hence our assumption that $\{f_n(z)\}$ does not converge uniformly to $f(z)$ inside G must be false. This completes the proof.

Vitali's theorem has some important implications concerning the analyticity of functions defined as complex integrals:

THEOREM 17.19. *Let G be a domain in the z-plane, and let L be a rectifiable curve in the w-plane. Suppose $\Phi(z, w)$ is a function of two complex variables with the following properties*:

1. *For every $w \in L$, $\Phi(z, w)$ is analytic in z on the domain G;*
2. *For every $z \in G$, $\Phi(z, w)$ is continuous in w on the curve L;*
3. *The family of functions of z*

$$\Phi(z, w), \quad w \in L,$$

is uniformly bounded inside G.

Then the integral

$$f(z) = \int_L \Phi(z, w)\, dw \tag{17.75}$$

is analytic on G. Moreover, if for every $z \in G$, the partial derivatives

$$\frac{\partial^m \Phi(z, w)}{\partial z^m} \qquad (m = 1, 2, \ldots) \tag{17.76}$$

are continuous in w on the curve L, then

$$f^{(m)}(z) = \int_L \frac{\partial^m \Phi(z, w)}{\partial z^m}\, dw \qquad (m = 1, 2, \ldots). \tag{17.77}$$

Proof. Suppose the equation of L is

$$w = \lambda(t) \qquad (\alpha \leqslant t \leqslant \beta).$$

We introduce a sequence of partitions

$$\mathscr{P}^{(n)} = \{t_0^{(n)}, t_1^{(n)}, \ldots, t_n^{(n)}\} \qquad (n = 1, 2, \ldots)$$

of the interval $[\alpha, \beta]$, where

$$\alpha = t_0^{(n)} < t_1^{(n)} < \cdots < t_{n-1}^{(n)} < t_n^{(n)} = \beta,$$

such that the norm of $\mathscr{P}^{(n)}$ approaches zero as $n \to \infty$ (cf. Sec. 60). Let

$$w_k^{(n)} = \lambda(t_k^{(n)}) \qquad (k = 0, 1, \ldots, n),$$

and form the sum

$$f_n(z) = \sum_{k=1}^{n} \Phi(z, w_k^{(n)})(w_k^{(n)} - w_{k-1}^{(n)}). \tag{17.78}$$

The sequence $\{f_n(z)\}$ is uniformly bounded inside G, since if F is any closed bounded subset of G, then

$$|f_n(z)| \leqslant M_F \sum_{k=1}^{n} |w_k^{(n)} - w_{k-1}^{(n)}| \leqslant M_F l \qquad (z \in F),$$

where l is the length of L and

$$M_F = \sup_{\substack{z \in F \\ w \in L}} |\Phi(z, w)|.$$

Therefore, by Theorem 17.17, $\{f_n(z)\}$ is compact on G. Moreover, since L is rectifiable and $\Phi(z, w)$ is continuous in w on the curve L for every $z \in G$, $\{f_n(z)\}$ converges on G to the function $f(z)$, given by (17.75). It follows from Vitali's theorem that $\{f_n(z)\}$ is uniformly convergent [to $f(z)$] inside G, and hence from Weierstrass' theorem for sequences of analytic functions (Theorem 15.8) that $f(z)$ is analytic on G. To prove (17.77), we note that Theorem 15.8, together with (17.78), also implies

$$f^{(m)}(z) = \lim_{n \to \infty} f_n^{(m)}(z) = \lim_{n \to \infty} \sum_{k=1}^{n} \frac{\partial^m \Phi(z, w_k^{(n)})}{\partial z^k} (w_k^{(n)} - w_{k-1}^{(n)})$$

$$= \int_L \frac{\partial^m \Phi(z, w)}{\partial z^m}\, dw.$$

Next we extend Theorem 17.19 to the important case where L is an unbounded curve (see p. 333).

THEOREM 17.20. *Let G be a domain in the z-plane, and let L be an unbounded curve in the w-plane, with equation*

$$w = \lambda(t) \qquad (\alpha \leqslant t < \beta),$$

where $\lambda(t)$ is continuous on $[\alpha, \beta)$ and approaches ∞ if and only if $t \to \beta -$.

Suppose every finite arc of L, i.e., every curve L_τ with equation

$$w = \lambda(t) \qquad (\alpha \leqslant t \leqslant \tau < \beta)$$

is rectifiable, and suppose $\Phi(z, w)$ is a function of two complex variables with the following properties:

1. *For every $w \in L$, $\Phi(z, w)$ is analytic in z on the domain G;*
2. *For every $z \in G$, $\Phi(z, w)$ is continuous in w on the curve L;*
3. *For every $\tau \in (\alpha, \beta)$, the family of functions*

$$\Phi(z, w), \quad w \in L_\tau$$

is uniformly bounded inside G;[23]

4. *For every $z \in G$, the limit*

$$\lim_{\tau \to \beta -} \int_{L_\tau} |\Phi(z, w)|\, ds = \int_L |\Phi(z, w)|\, ds \qquad (ds = |dw|) \quad (17.79)$$

exists, i.e., the integral

$$\lim_{\tau \to \beta -} \int_{L_\tau} \Phi(z, w)\, dw = \int_L \Phi(z, w)\, dw$$

is absolutely convergent (and hence exists);

5. *The function* (17.79) *is bounded inside G.*

Then the integral

$$f(z) = \int_L \Phi(z, w)\, dw \qquad (17.80)$$

is analytic on G. Moreover, if for every $z \in G$, the partial derivatives (17.76) *are continuous in w on the curve L, then $f^{(m)}(z)$ is given by* (17.77).

Proof. Using Theorem 17.19 and properties 1, 2 and 3, we see that every function

$$f_\tau(z) = \int_{L_\tau} \Phi(z, w)\, dw$$

is analytic on G. The family $\{f_\tau(z)\}$ is uniformly bounded inside G, since

$$|f_\tau(z)| \leqslant \int_{L_\tau} |\Phi(z, w)|\, ds \leqslant \int_L |\Phi(z, w)|\, ds$$

(cf. properties 4 and 5). Therefore, by Theorem 17.17, the family $\{f_\tau(z)\}$, or any sequence $\{f_{\tau_n}(z)\}$ where $\tau_n \to \beta -$ as $n \to \infty$, is compact. Applying Vitali's theorem, we see that $\{f_{\tau_n}(z)\}$, which converges on G to the function

---

[23] In other words, for every closed bounded subset $F \subset G$ and every $\tau \in (\alpha, \beta)$ there exists a number $M(F, \tau) > 0$ such that

$$|\Phi(z, w)| \leqslant M(F, \tau)$$

for all $z \in F$ and $w \in L_\tau$.

$f(z)$, given by (17.80), is uniformly convergent inside G. Therefore it follows from Weierstrass' theorem for families of analytic functions (Theorem 15.12) that $f(z)$ is analytic on G. To prove (17.77) for the unbounded curve L, we note that Theorem 15.12, together with the validity of (17.77) for every finite arc L_τ, implies

$$f^{(m)}(z) = \lim_{\tau \to \beta -} f_\tau^{(m)}(z) = \lim_{\tau \to \beta -} \int_{L_\tau} \frac{\partial^m \Phi(z, w)}{\partial z^m}\, dw = \int_L \frac{\partial^m \Phi(z, w)}{\partial z^m}\, dw \qquad (m = 1, 2, \ldots).$$

Finally we generalize Theorem 17.19 to the case where $\Phi(z, w)$ can become infinite at an end point of L:

THEOREM 17.21. *Let G be a domain in the z-plane, and let L be a rectifiable curve in the w-plane, with equation*

$$w = \lambda(t) \qquad (\alpha \leqslant t \leqslant \beta).$$

Let L_τ be the arc of L with equation

$$w = \lambda(t) \qquad (\alpha \leqslant t \leqslant \tau < \beta),$$

and suppose $\Phi(z, w)$ is a function of two complex variables with the following properties:

1. *For some $z_0 \in G$,*
$$\lim_{t \to \beta -} \Phi[z_0, \lambda(t)] = \infty;$$
2. *For every $w \in L_\tau$ and $\tau \in (\alpha, \beta)$, $\Phi(z, w)$ is analytic in z on the domain G;*
3. *For every $z \in G$ and $\tau \in (\alpha, \beta)$, $\Phi(z, w)$ is continuous in w on the curve L_τ;*
4. *For every $\tau \in (\alpha, \beta)$, the family of functions*
$$\Phi(z, w), \quad w \in L_\tau$$
is uniformly bounded inside G;
5. *For every $z \in G$, the limit*
$$\lim_{\tau \to \beta -} \int_{L_\tau} |\Phi(z, w)|\, ds = \int_L |\Phi(z, w)|\, ds \qquad (ds = |dw|)$$
exists and is bounded inside G.

Then the integral

$$(z) = \int_L \Phi(z, w)\, dw$$

is analytic on G. Moreover, if for every $z \in G$ and $\tau \in (\alpha, \beta)$, the partial derivatives (17.76) *are continuous in w on the curve L_τ, then $f^{(m)}(z)$ is given by* (17.77).

Proof. Word for word, the proof is the same as that of Theorem 17.20, except that now it is the unboundedness of $\Phi(z, w)$ which is at issue, rather than the unboundedness of L.

Example. Let G be the whole z-plane, let L be the nonnegative real axis $0 \leqslant t < \infty$, and let

$$\Phi(z, t) = e^{-zt}\varphi(t),$$

where $\varphi(t)$ is real and continuous on L. Then the integral (17.80) becomes

$$f(z) = \int_0^\infty e^{-zt}\varphi(t)\, dt, \tag{17.81}$$

and is known as the *Laplace transform* of $\varphi(t)$. Obviously, $\Phi(z, t)$ satisfies properties 1 and 2 of Theorem 17.20, and the fact that $\Phi(z, t)$ satisfies property 3 as well is an easy consequence of the relation

$$|\Phi(z, t)| = e^{-xt}|\varphi(t)|.$$

Suppose, in addition, that there are positive numbers α and C such that

$$|\varphi(t)| < \alpha e^{Ct} \quad \text{if} \quad t > T(\alpha, C). \tag{17.82}$$

Then

$$|\Phi(z, t)| < \alpha e^{-t(x-C)} \quad \text{if} \quad t > T(\alpha, C),$$

and the integral

$$\int_0^\infty |\Phi(z, t)|\, dt \tag{17.83}$$

converges for all z belonging to the half-plane $x = \operatorname{Re} z > C$. Moreover, (17.83) is bounded in every half-plane $\operatorname{Re} z \geqslant C + \varepsilon$ ($\varepsilon > 0$), since if $x \geqslant C + \varepsilon$, then

$$\int_0^\infty |\Phi(z, t)|\, dt < A + \alpha \int_0^\infty e^{-\varepsilon t}\, dt = A + \frac{\alpha}{\varepsilon},$$

where A is any positive number exceeding

$$\int_0^{T(\alpha,C)} |\Phi(z, t)|\, dt,$$

e.g.,

$$A = \int_0^{T(\alpha,C)} e^{-Ct}\, |\varphi(t)|\, dt.$$

Therefore, if the condition (17.82) holds, $\Phi(z, t)$ has properties 4 and 5 of Theorem 17.20, and hence the function $f(z)$, defined by the integral (17.81), is analytic in the half-plane $\operatorname{Re} z > C$. Finally, it follows from (17.77) that

$$f^{(m)}(z) = \int_0^\infty (-t)^m e^{-zt}\varphi(t)\, dt \qquad (m = 1, 2, \ldots)$$

in the same half-plane.

PROBLEMS

17.1. Give an alternative derivation of the addition theorem for $\cos z$ and $\sin z$ (see Sec. 40).

Hint. Use Theorem 17.1, regarding the addition theorems as known for real arguments.

17.2. Does there exist a function which is analytic at the point $z = 0$ and takes the following values at the points $z = 1/n$ $(n = 1, 2, \ldots)$:

a) $0, 1, 0, 1, 0, 1, \ldots 0, 1, \ldots$;

b) $0, \frac{1}{2}, 0, \frac{1}{4}, 0, \frac{1}{6}, \ldots, 0, \frac{1}{2k}, \ldots$;

c) $\frac{1}{2}, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{6}, \frac{1}{6}, \ldots, \frac{1}{2k}, \frac{1}{2k}, \ldots$;

d) $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{6}{7}, \ldots, \frac{n}{n+1}, \ldots$?

17.3. Does there exist a function which is analytic at the point $z = 0$ and satisfies the following condition for every positive integer n:

a) $f(1/n) = f(-1/n) = 1/n^2$; b) $f(1/n) = f(-1/n) = 1/n^3$?

Ans. a) Yes; b) no.

17.4. Prove that every set of measure zero in the sense of the definition on p. 310 is measurable and has measure zero in the sense of the definition on p. 370.

17.5. Prove that if $f(z)$ and $g(z)$ are analytic at a point z_0 and have zeros at z_0 of orders p and q, respectively, then $f(z)g(z)$ is analytic at z_0 and has a zero at z_0 of order pq.

17.6. Find the order of the zero at $z = 0$ of the function

a) $z^2(e^{z^2} - 1)$; b) $6 \sin z^3 + z^3(z^6 - 6)$; c) $e^{\sin z} - e^{\tan z}$.

Ans. b) 15.

17.7. Find the zeros (possibly including ∞), together with their orders, of each of the following functions:

a) $z^2 + 9$; b) $\dfrac{z^2 + 9}{z^4}$; c) $z \sin z$; d) $(1 - e^z)(z^2 - 4)^3$;

e) $1 - \cos z$; f) $\dfrac{(z^2 - \pi^2)^2 \sin z}{z^7}$; g) $\dfrac{1 - \cot z}{z}$; h) $e^{\tan z}$;

i) $\sin^3 z$; j) $\dfrac{\sin^3 z}{z}$; k) $\sin z^3$; l) $\cos^3 z$; m) $\cos z^3$; n) $(\sqrt{z} - 2)^3$;

o) $(1 - \sqrt{2 - 2\cos z})^2$.

17.8. The function

$$\sin \frac{1}{1 - z},$$

although obviously not a constant, has an infinite sequence of zeros converging to the point $z = 1$. Why does this not contradict the considerations of Sec. 82?

17.9. Let G be a domain, and let $F_1, F_2, \ldots, F_n, \ldots$ be the increasing sequence of compact sets approximating G, constructed on p. 373. Why does the fact that every F_n is closed while G is open represent no contradiction? Does the construction apply in the case where G is an arbitrary open set (not necessarily a domain)? Must every F_n be a continuum if G is a domain?

17.10. Find an increasing sequence of compact sets approximating

a) The interval $(-1, 2)$; b) The interval $(-2, +\infty)$;
c) The interior of the ellipse $(x^2/a^2) + (y^2/b^2) = 1$;
d) The set of points $z = x + iy$ such that $y^2 > |x|$;
e) The union of the interior of the circle $|z| = 2$ and the exterior of the circle $|z| = 4$.

17.11. Prove that the infinite product

$$\prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2\pi^2}\right)$$

represents an entire function whose only zeros are at the points $\pm\pi, \pm 2\pi, \ldots$.

Comment. The entire function in question is $\sin z/z$.

17.12. Prove that if we can find a complex number ζ such that the nonzero terms $a_n\zeta^n$ in the power series

$$f(z) = \sum_{n=0}^{\infty} a_n z^n$$

all have the same argument, then

$$M(|\zeta|) = |f(\zeta)|,$$

where $M(\rho)$ is the function

$$M(\rho) = \max_{|z|=\rho} |f(z)|,$$

defined for all ρ less than the radius of convergence of the power series.

17.13. Use the result of the preceding problem to find $M(\rho)$ for the following entire functions:

a) e^z; b) $\cos z$; c) $\sin z$; d) $\dfrac{\sin \sqrt{z}}{\sqrt{z}}$.

17.14. Give an alternative proof of the fundamental theorem of algebra (Theorem 17.7), based on the use of Cauchy's integral theorem.

Hint. If the theorem is false, then

$$f(z) = \frac{1}{P_n(z)\bar{P}_n(z)}$$

is an entire function, where $\bar{P}_n(z)$ is the polynomial obtained by replacing all the coefficients of $P_n(z)$ by their complex conjugates. Integrate $f(z)$ along the

closed contour L shown in Figure 17.9. According to Cauchy's integral theorem,

$$0 = \int_L f(z)\,dz = \int_{-R}^{R} f(x)\,dx + i\int_0^{\pi} f(Re^{i\theta})\,Re^{i\theta}\,d\theta.$$

Show that this is a contradiction.

17.15. Give an example of a lemniscate $|P_n(z)| = r^n$ whose n components for small r remain disjoint for $r < 1$ and coalesce into a single component for $r = 1$.

17.16. Generalize Schwarz' lemma (Theorem 17.8) to the case where $f(\alpha) = 0$ ($|\alpha| < R$), instead of $f(0) = 0$.

Hint. Cf. Example 2, p. 182

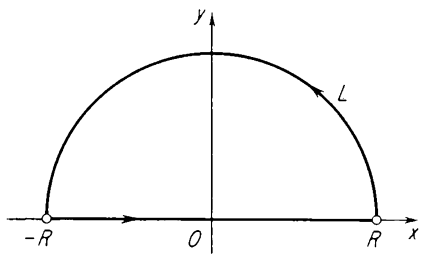

FIGURE 17.9

17.17. Show that the domain D figuring in the proof of Theorem 17.9 is simply connected.

17.18. Prove that every point of the unit circle $|z| = 1$ is a singular point of the power series

$$\sum_{n=0}^{\infty} z^{n!}.$$

17.19. Prove that the series (17.31) converges uniformly to the kth derivative of the series (17.30) on the closed unit disk $\bar{K}$: $|z| \leqslant 1$, where, if $|\zeta| = 1$, $f^{(k)}(\zeta)$ means

$$\lim_{\substack{z \to \zeta \\ |z|<1}} \frac{f^{(k-1)}(z) - f^{(k-1)}(\zeta)}{z - \zeta}.$$

Hint. Let $s(z)$ denote the sum of (17.31) for $k = 1$. Then term-by-term integration of (17.31) gives

$$\int_{z_0}^{z} s(\zeta)\,d\zeta = f(z) - f(z_0),$$

where $z_0 \in \bar{K}$ is fixed and $z \in \bar{K}$ is variable [cf. Problem 15.6]. It follows that $s(z) = f'(z)$ [cf. Theorem 13.5]. Complete the proof by induction.

Comment. Note that Theorem 15.6 implies the result only for $|z| < 1$.

17.20. Let $f(z)$ be an analytic function defined on a domain G, with boundary Γ. Suppose the point $\zeta \in \Gamma$ and the set $E \subset G$ with limit point ζ are such that

$$\lim_{\substack{z \to \zeta \\ z \in E}} f(z) = \infty.$$

Prove that ζ is a singular point of some element $\{K, f(z)\}$, where $K \subset G$ and ζ

belongs to the boundary of K. Prove that if $E = G$, then ζ is a singular point of every element $\{K, f(z)\}$ with ζ on the boundary of K.

17.21. Let G and $f(z)$ be the same as in Example 2, p. 393. Prove that if z_0 belongs to the upper half-plane, then the radius of convergence of the Taylor series of $f(z)$ at the point z_0 equals the distance from z_0 to the nearer of the two points 0, 1 (or to either, if both lie at the same distance from z_0). Prove that if z_0^* belongs to the lower half-plane, then the radius of convergence of the Taylor series of $f(z)$ at the point z_0^* equals the distance from z_0^* to the origin (in this case, the distance from z_0^* to the point 1 plays no role at all).

17.22. Let $\{a_n\}$ be a sequence of positive numbers which decreases monotonically to zero, and suppose the power series

$$\sum_{n=0}^{\infty} a_n z^n$$

has radius of convergence 1. Prove that the series converges at every point of its circle of convergence, with the possible exception of the point $z = 1$.

Hint. Cf. formula (3.16) and Example 1, p. 399.

17.23. Let

$$f(z) = \sum_{n=0}^{\infty} a_n z^n$$

be a power series with radius of convergence 1, and suppose that

$$\sum_{n=0}^{\infty} a_n e^{in\theta}$$

converges. Prove that

$$\lim_{z \to e^{i\theta}} f(z) = \sum_{n=0}^{\infty} a_n e^{in\theta},$$

where $z \to e^{i\theta}$ in the same way as $z \to 1$ in the statement of Theorem 17.14.

17.24. Verify the following Fourier expansions:

a) $$-\ln\left(2 \sin \frac{\theta}{2}\right) = \sum_{n=1}^{\infty} \frac{\cos n\theta}{n} \qquad (0 < \theta < 2\pi);$$

b) $$\frac{\pi - \theta}{2} = \sum_{n=1}^{\infty} \frac{\sin n\theta}{n} \qquad (0 < \theta < 2\pi);$$

c) $$\frac{1}{2} \ln\left(\cot \frac{\theta}{2}\right) = \sum_{n=0}^{\infty} \frac{\cos (2n+1)\theta}{2n+1} \qquad (0 < \theta < \pi);$$

d) $$\frac{\pi}{4} = \sum_{n=0}^{\infty} \frac{\sin (2n+1)\theta}{2n+1} \qquad (0 < \theta < \pi).$$

Hint. Cf. Theorem 17.14 and Problem 16.13.

17.25. Prove that the condition (17.46) in Tauber's theorem can be replaced by the condition (17.47).

Hint. If

$$A_n = a_1 + 2a_2 + \cdots + na_n \qquad (n > 0), \quad A_0 = 0,$$

then

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + (1 - x) \sum_{n=1}^{\infty} \frac{A_n}{n} x^n + \sum_{n=1}^{\infty} \frac{A_n}{n(n+1)} x^{n+1}.$$

Prove that the second term on the right approaches 0 as $x \to 1-$, and then apply Theorem 17.15 to the series

$$\sum_{n=1}^{\infty} \frac{A_n}{n(n+1)}.$$

Finally transform back to the coefficients a_n.

17.26. Prove that the condition (17.46) in Tauber's theorem can be replaced by the condition

$$\lim_{n \to \infty} a_n = 0$$

if all the a_n are nonnegative.

17.27. Prove that the series

$$\sum_{n=0}^{\infty} \frac{z^{2n+1}}{2n+1}$$

is uniformly convergent on every arc $\sigma: z = e^{i\theta}$ $(\alpha \leqslant \theta \leqslant \beta)$ which does not contain either of the points $z = \pm 1$.

Hint. Cf. Problem 16.13.

17.28. Let G be a domain, and let $\mathscr{F}$ be a family of analytic functions on G which is uniformly bounded inside G. Prove that the family consisting of the derivatives of the functions in $\mathscr{F}$ is also uniformly bounded inside G.

Hint. Given a closed bounded subset $F \subset G$, construct a closed Jordan curve L such that $L \subset G$, $F \subset I(L)$, and then use formula (14.20).

17.29. Let G be a domain, and let $\mathscr{F}$ be a family of analytic functions on G which is uniformly bounded inside G. Prove that $\mathscr{F}$ is *equicontinuous* inside G, i.e., prove that given any $\varepsilon > 0$ and any closed bounded subset $F \subset G$, there exists a number $\delta(\varepsilon) > 0$ such that $|f(z) - f(z')| < \varepsilon$ for all $f(z) \in \mathscr{F}$ if z, z' are any two points in F satisfying the inequality $|z - z'| < \delta(\varepsilon)$.

Hint. Use the result of the preceding problem.

17.30. Verify that

$$\frac{\cos \theta}{1 - z} \qquad (-\infty < \theta < \infty)$$

is a compact family of analytic functions on the unit disk $K: |z| < 1$.

17.31. Show that both of the integrals

$$\int_0^\infty e^{-zt^2}\, dt, \qquad \int_0^\infty e^{-t} t^{z-1}\, dt$$

define analytic functions for $\operatorname{Re} z > 0$.

Comment. The second integral is the *gamma function* $\Gamma(z)$ [cf. Problem 13.11].

17.32. Prove that

$$\ln z = \int_0^\infty \frac{e^{-t} - e^{-zt}}{t}\, dt \qquad (\operatorname{Re} z > 0).$$

Hint. Use Theorem 17.20 to verify that the integral defines an analytic function for $\operatorname{Re} z > 0$. Then expand the integrand as a power series and integrate term by term. Use formula (16.58) to identify the result as $\ln z$ on the disk $|z - 1| < 1$. Finally, use Theorem 17.1 to extend the result to the half-plane $\operatorname{Re} z > 0$.

17.33. Verify that Theorem 14.5 is a special case of Theorem 17.19. Verify that the example on pp. 333–334 is a special case of Theorem 17.20.

17.34. Investigate the Laplace transform

$$f(z) = \int_0^\infty e^{-zt} \varphi(t)\, dt$$

of the function

a) $\varphi(t) = 1$; b) $\varphi(t) = t^n$; c) $\varphi(t) = e^{-t^2}$; d) $\varphi(t) = e^{t^2}$.

17.35. Prove that with the hypotheses of Theorems 17.19 through 17.21, indefinite integration behind the integral sign is permissible, i.e.,

$$\int_{z_0}^{z} f(\zeta)\, d\zeta = \int_L \left[\int_{z_0}^{z} \Phi(\zeta, w)\, d\zeta\right] dw.$$

Hint. Cf. Problems 15.6 and 15.7.

18

METHODS FOR EXPANDING FUNCTIONS IN TAYLOR SERIES

88. The Taylor Series of the Sum of a Series of Analytic Functions

In principle, the problem of finding the Taylor series expansion of an analytic function $f(z)$ in a neighborhood of a point z_0 is solved by using the formula

$$a_n = \frac{f^{(n)}(z_0)}{n!} \qquad (n = 0, 1, 2, \ldots) \tag{18.1}$$

to find the coefficients of the series

$$f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n. \tag{18.2}$$

However, it often turns out to be a formidable task to calculate the coefficients (18.1) directly. Fortunately, in many practical problems, the Taylor series (18.2) can readily be deduced by starting from other series expansions with known coefficients.

THEOREM 18.1 *Suppose the series*

$$f(z) = \sum_{k=1}^{\infty} f_k(z) \tag{18.3}$$

is uniformly convergent on every compact subset of the disk $K: |z - z_0| < R$, *and suppose every function* $f_k(z)$ *is analytic on* K. *Then* $f(z)$ *has the*

Taylor series expansion (18.2) *on* K, *where the coefficients* (18.1) *are related to the derivatives of the original series* (18.3) *by the formula*

$$a_n = \sum_{k=1}^{\infty} \frac{f_k^{(n)}(z_0)}{n!} \qquad (n = 0, 1, 2, \ldots).$$

Proof. According to Theorem 15.6, $f(z)$ is analytic on K, and hence, according to Theorem 16.7, $f(z)$ has the Taylor series expansion (18.2) on K, with coefficients (18.1). The rest of the proof follows at once by combining (18.1) and the formula

$$f^{(n)}(z) = \sum_{k=1}^{\infty} f_k^{(n)}(z) \qquad (n = 0, 1, 2, \ldots)$$

[cf. (15.17)].

Remark. Since each of the functions $f_k(z)$ in the right-hand side of (18.3) has its own Taylor series expansion on K, i.e.,

$$f_k(z) = \sum_{l=0}^{\infty} a_{kl}(z - z_0)^l \qquad (k = 1, 2, \ldots),$$

where

$$a_{kl} = \frac{f_k^{(l)}(z_0)}{l!} \qquad (l = 0, 1, 2, \ldots),$$

it follows from Theorem 18.1 that

$$a_n = \sum_{k=1}^{\infty} a_{kn}.$$

In other words, the coefficient of a given power of $z - z_0$ in the right-hand side of (18.2) is obtained by adding all the coefficients of the same power of $z - z_0$ in the separate Taylor series expansions of $f_k(z)$, $k = 1, 2, \ldots$.

Example 1. If

$$f_k(z) = \frac{z^k}{1 - z^k},$$

(18.3) becomes

$$f(z) = \sum_{k=1}^{\infty} \frac{z^k}{1 - z^k}. \tag{18.4}$$

Every term $f_k(z)$ is analytic on the unit disk $K: |z| < 1$, and the series (18.4) is uniformly convergent on every compact subset of K. In fact, let $F \subset K$ be compact, and let $\delta > 0$ be the distance between F and the unit circle $|z| = 1$. Then $|z| \leqslant 1 - \delta = \rho < 1$ for all $z \in F$, and hence

$$\left|\frac{z^k}{1 - z^k}\right| \leqslant \frac{\rho^k}{1 - \rho^k} \leqslant \frac{\rho^k}{1 - \rho}.$$

Since the series

$$\sum_{k=1}^{\infty} \frac{\rho^k}{1 - \rho}$$

is convergent (being a geometric series with ratio $\rho < 1$), it follows from Weierstrass' M-test (Theorem 15.2) that the series (18.4) is uniformly convergent on F, as asserted.

According to Theorem 18.1, to find the coefficient of z^n in the Taylor series of the analytic function $f(z)$, we need only add all the coefficients of z^n in the separate Taylor series

$$f_k(z) = \frac{z^k}{1 - z^k} = z^k + z^{2k} + \cdots.$$

In each such expansion, the coefficient of z^n equals 1 if n is divisible by k, and 0 otherwise. Therefore the coefficient of z^n in the expansion of $f(z)$ equals the number of positive integers which divide n (including 1 and n). Denoting this number by $\tau(n)$,[1] we have

$$f(z) = \sum_{n=1}^{\infty} \tau(n) z^n.$$

Example 2. If

$$f_k(z) = \frac{2z}{z^2 - k^2\pi^2},$$

(18.3) becomes

$$f(z) = \sum_{k=1}^{\infty} \frac{2z}{z^2 - k^2\pi^2}. \tag{18.5}$$

The function $f_k(z)$ is analytic in the whole complex plane, except at the points $z = \pm k\pi$, at which $f_k(z)$ becomes infinite. Therefore every term $f_k(z)$ is analytic on the whole disk $K: |z| < \pi$. Moreover, the series (18.5) is uniformly convergent on every compact subset of K. In fact, if $|z| \leqslant \rho$, where $0 < \rho < \pi$, then

$$\left|\frac{2z}{z^2 - k^2\pi^2}\right| \leqslant \frac{2\rho}{k^2\pi^2 - \rho^2} = \frac{2\rho}{k^2\pi^2} \frac{1}{1 - \frac{\rho^2}{k^2\pi^2}} \leqslant \frac{2\rho}{k^2\pi^2} \frac{1}{1 - \frac{\rho^2}{\pi^2}},$$

where the right-hand side is the general term of a convergent series.[2] Therefore, by Weierstrass' M-test, the series (18.5) is uniformly convergent on every closed disk $|z| \leqslant \rho < \pi$, and hence on every compact subset of K.

---

[1] Thus $\tau(1) = 1$, $\tau(2) = 2$, $\tau(3) = 2$, $\tau(4) = 3$, $\tau(5) = 2, \ldots$

[2] The series

$$\sum_{k=1}^{\infty} \frac{1}{k^2}$$

is convergent, and in fact equals $\pi^2/6$.

Thus $f(z)$ is analytic on K, and the coefficient of z^n in the Taylor series of $f(z)$ equals the sum of the coefficients of z^n in the separate Taylor series

$$f_k(z) = \frac{2z}{z^2 - k^2\pi^2} = -\frac{2z}{k^2\pi^2}\frac{1}{1 - \dfrac{z^2}{k^2\pi^2}}$$
$$= -\frac{2z}{k^2\pi^2} - \frac{2z^3}{k^4\pi^4} - \cdots - \frac{2z^{2m-1}}{k^{2m}\pi^{2m}} - \cdots.$$

Here the coefficient of z equals 0 if n is even, and $-2/k^{2m}\pi^{2m}$ if $n = 2m - 1$ is odd. Therefore the coefficient of z^n in the expansion of $f(z)$ equals 0 if n is even, and

$$-2\sum_{k=1}^{\infty}\frac{1}{(k\pi)^{2m}}$$

if $n = 2m - 1$ is odd. It follows that

$$f(z) = -2\sum_{m=1}^{\infty}\left[\sum_{k=1}^{\infty}\frac{1}{(k\pi)^{2m}}\right]z^{2m-1}, \tag{18.6}$$

which must converge for $|z| < \pi$, since $f(z)$ is analytic on K. At the point $z = \pi$, each term of the series satisfies the inequality

$$\sum_{k=1}^{\infty}\frac{1}{(k\pi)^{2m}}\pi^{2m-1} = \frac{1}{\pi}\sum_{k=1}^{\infty}\frac{1}{k^{2m}} > \frac{1}{\pi},$$

and hence the series (18.6) diverges for $z = \pi$ (since the necessary condition for convergence is not fulfilled), i.e., the series (18.6) has radius of convergence π.

89. The Taylor Series of a Composite Function

Given a Taylor series

$$f(z) = a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots \tag{18.7}$$

at the point z_0 in the z-plane, and a second Taylor series

$$\varphi(w) = A_0 + A_1(w - w_0) + \cdots + A_n(w - w_0)^n + \cdots \tag{18.8}$$

at the point $w_0 = f(z_0) = a_0$ in the w-plane, suppose we replace w by $f(z)$ in (18.8). Then we obtain a new series expansion

$$F(z) = A_0 + A_1[f(z) - a_0] + \cdots + A_n[f(z) - a_0]^n + \cdots, \tag{18.9}$$

corresponding to the *composite function* $F(z) = \varphi[f(z)]$.[3] Suppose the algebraic operations called for in the right-hand side of (18.9) are carried out formally: This involves raising $f(z) - a_0$ to various powers after replacing $f(z)$ by the series (18.7), and then grouping together and adding all the coefficients of each power $(z - z_0)^n$, $n = 0, 1, 2, \ldots$ in the expression so obtained. The result is a particular Taylor series of the form

$$F(z) = b_0 + b_1(z - z_0) + \cdots + b_n(z - z_0)^n + \cdots, \tag{18.10}$$

which is said to be obtained by *formal substitution* of the series (18.7) into the series (18.8).

THEOREM 18.2. *If $f(z)$ is analytic at the point $z = z_0$, and if $\varphi(w)$ is analytic at the point $w_0 = f(z_0)$, then the composite function $F(z) = \varphi[f(z)]$ is analytic at the point $z = z_0$. Moreover, the Taylor series of $F(z)$ is obtained by formal substitution of the Taylor series of $w = f(z)$ into the Taylor series of $\varphi(w)$.*

Proof. Let (18.7) be the Taylor series of $f(z)$, valid for $|z - z_0| < r$, and let (18.8) be the Taylor series of $\varphi(w)$, valid for $|w - w_0| < R$, where $w_0 = f(z_0) = a_0$. (By hypothesis, such positive numbers r and R exist.) Since $f(z) \to f(z_0)$ as $z \to z_0$, there is a number r' $(0 < r' \leqslant r)$ such that $|f(z) - w_0| < R' < R$ if $|z - z_0| < r'$. Clearly, the expansion (18.9) holds for all $|z - z_0| < r'$. This expansion is of the form

$$F(z) = \sum_{k=0}^{\infty} F_k(z), \tag{18.11}$$

where every function

$$F_k(z) = A_k[f(z) - a_0]^k \qquad (k = 0, 1, 2, \ldots)$$

is analytic on the disk K: $|z - z_0| < r'$. Moreover, according to Theorem 16.3, the series (18.11), regarded as the power series (18.8) in the w-plane, is uniformly convergent on K, and hence, *a fortiori*, on every compact subset of K. It follows from Theorem 18.1 and the remark on p. 430 that $F(z)$ is analytic on K (in particular at z_0), and that the coefficient of a given power $(z - z_0)^n$ in the series (18.11) is obtained by adding all the coefficients of the same power $(z - z_0)^n$ in the separate Taylor series of $F_k(z)$, $k = 0, 1, 2, \ldots$. All that remains to be proved is that the Taylor series of $F_k(z)$ is obtained by raising the Taylor series of $f(z) - a_0$ to the kth power and multiplying the result by A_k. But this is an immediate consequence of Theorem 3.7 and the fact that the Taylor series of $f(z) - a_0$ is absolutely convergent on K.

---

[3] As it stands, (18.9) is not a Taylor series in the z-plane. Note that $F(z_0) = \varphi(w_0) = A_0$, as required. Thus b_0 must equal A_0 in (18.10).

Remark. The fact that $F(z)$ is differentiable for $|z - z_0| < r'$, and hence analytic at z_0, also follows at once from the rule for differentiation of a composite function (see Rule 4, p. 108).

Example 1. Let

$$F(z) = +\sqrt{\cos z}, \tag{18.12}$$

where $+\sqrt{\cos z}$ denotes the single-valued branch of the double-valued function $\sqrt{\cos z}$ which is analytic at $z = 0$ and takes the value $+1$ at $z = 0$. With this understanding, we shall henceforth omit the plus sign in (18.12). To apply Theorem 18.2, we write (18.12) in the form

$$F(z) = [1 - (1 - \cos z)]^{1/2}.$$

Then

$$w = f(z) = 1 - \cos z = \frac{z^2}{2!} - \frac{z^4}{4!} + \frac{z^6}{6!} - \cdots \qquad (|z| < \infty)$$

and

$$\varphi(w) = (1 - w)^{1/2} = 1 - \frac{w}{2} - \frac{w^2}{8} - \frac{w^3}{16} - \cdots \qquad (|w| < 1)$$

[cf. (16.60)], so that

$$F(z) = \varphi[f(z)] = 1 - \frac{1}{2}\left(\frac{z^2}{2} - \frac{z^4}{24} + \frac{z^6}{720} - \cdots\right) - \frac{1}{8}\left(\frac{z^2}{2} - \frac{z^4}{24} + \cdots\right)^2 - \frac{1}{16}\left(\frac{z^2}{2} - \cdots\right)^3 - \cdots,$$

where the omitted terms contribute only to the coefficients of z^8, $z^{10}, \ldots$ in the Taylor series of $F(z)$. Thus, writing

$$\left(\frac{z^2}{2} - \frac{z^4}{24} + \cdots\right)^2 = \frac{z^4}{4} - \frac{z^6}{24} + \cdots, \qquad \left(\frac{z^2}{2} - \cdots\right)^3 = \frac{z^6}{8} - \cdots,$$

we find that

$$\begin{aligned} f(z) &= 1 - \frac{1}{2}\left(\frac{z^2}{2} - \frac{z^4}{24} + \frac{z^6}{720} - \cdots\right) - \frac{1}{8}\left(\frac{z^4}{4} - \frac{z^6}{24} + \cdots\right) - \frac{1}{16}\left(\frac{z^6}{8} - \cdots\right) \\ &= 1 - \frac{z^2}{4} - \frac{z^4}{96} - \frac{19z^6}{5760} - \cdots. \end{aligned} \tag{18.13}$$

The series (18.13) converges on the disk $K: |z| < \pi/2$, since $F(z) = \sqrt{\cos z}$ is analytic on K, with derivative

$$F'(z) = -\frac{\sin z}{2\sqrt{\cos z}} \qquad (z \in K).$$

Moreover, $\pi/2$ is the radius of convergence of (18.13), since the element $\{K, F(z)\}$ has a singular point at $z = \pi/2$ (why?).

Example 2. Consider the function

$$F(z) = \exp \frac{1}{1 - z},$$

which can also be written in the form

$$F(z) = e \exp \frac{z}{1 - z}$$

(see Problem 18.4). In this case,

$$w = f(z) = \frac{z}{1 - z} = z + z^2 + z^3 + \cdots \qquad (|z| < 1),$$

and

$$\varphi(w) = e^{1+w} = e\left(1 + \frac{w}{1!} + \frac{w^2}{2!} + \cdots\right) \qquad (|w| < \infty).$$

Formal substitution of the second series into the first gives

$$F(z) = e\left[1 + \frac{(z + z^2 + z^3 + \cdots)}{1!} + \frac{(z + z^2 + z^3 + \cdots)^2}{2!} + \cdots\right]. \quad (18.14)$$

To simplify our calculations, we note that if $|z| < 1$, then

$$(z + z^2 + z^3 + \cdots)^k = \left(\frac{z}{1 - z}\right)^k = z^k(1 - z)^{-k}$$
$$= z^k\left[1 + \frac{k}{1!}z + \frac{k(k + 1)}{2!}z^2 + \cdots + \frac{k(k + 1)\cdots(k + l - 1)}{l!}z^l + \cdots\right],$$

or

$$(z + z^2 + z^3 + \cdots)^k = z^k\left[1 + \frac{k}{1!}z + \frac{k(k + 1)}{2!}z^2 + \cdots + \frac{(k + l - 1)(k + l - 2)\cdots(l + 1)}{(k - 1)!}z^l + \cdots\right]$$
$$= \sum_{l=0}^{\infty} \binom{l + k - 1}{k - 1} z^{l+k},$$

where

$$\binom{\alpha}{n} = \frac{\alpha(\alpha - 1)\cdots(\alpha - n + 1)}{n!}$$

is the binomial coefficient (see p. 400), defined for nonnegative integers n and arbitrary complex α (in the present case, α is also a nonnegative integer and

$\alpha \geqslant n$). Therefore (18.14) becomes

$$F(z) = e\Big\{1 + \frac{1}{1!}\sum_{l=0}^{\infty} z^{l+1} + \frac{1}{2!}\sum_{l=0}^{\infty}(l+1)z^{l+2} + \frac{1}{3!}\sum_{l=0}^{\infty}\frac{(l+1)(l+2)}{2}z^{l+3}$$
$$+ \frac{1}{4!}\sum_{l=0}^{\infty}\frac{(l+1)(l+2)(l+3)}{6}z^{l+4} + \cdots$$
$$+ \frac{1}{k!}\sum_{l=0}^{\infty}\binom{l+k-1}{k-1}z^{l+k} + \cdots\Big\}$$
$$= e\Big\{1 + z + \Big[\frac{1}{1!} + \frac{1}{2!}\Big]z^2 + \Big[\frac{1}{1!} + \frac{2}{2!} + \frac{1}{3!}\Big]z^3$$
$$+ \Big[\frac{1}{1!} + \frac{3}{2!} + \frac{3}{3!} + \frac{1}{4!}\Big]z^4$$
$$+ \cdots + \Big[\frac{1}{1!} + \binom{n-1}{1}\frac{1}{2!} + \binom{n-1}{2}\frac{1}{3!} + \cdots$$
$$+ \binom{n-1}{n-2}\frac{1}{(n-1)!} + \frac{1}{n!}\Big]z^n + \cdots\Big\}. \qquad (18.15)$$

The series (18.15) converges on the disk $K: |z| < 1$, since

$$F(z) = \exp\,[1/(1-z)]$$

is analytic on K. Moreover, the element $\{K, F(z)\}$ has a singular point at $z = 1$, since

$$\lim_{x \to 1-} \exp\frac{1}{1-x} = \infty,$$

and hence (18.15) has radius of convergence 1.

90. Division of Power Series

Let

$$a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots \qquad (18.16)$$

be a power series with radius of convergence $r > 0$, and let

$$b_0 + b_1(z - z_0) + \cdots + b_n(z - z_0)^n + \cdots \qquad (b_0 \neq 0) \quad (18.17)$$

be a second power series (in the same variable $z - z_0$) with radius of convergence $\rho > 0$. Since we assume $b_0 \neq 0$, there is a disk $K: |z - z_0| < R$ such that both series converge on K, while the second series has no zeros in K. In fact, we need only choose

$$R = \min\,(r, \rho, |z_1 - z_0|),$$

where z_1 is the zero of (18.17) which lies closest to z_0.[4] It follows from the

[4] If there is no unique z_1, we choose any of the (finitely many) zeros which lie at the same distance from z_0. If (18.17) has no zeros at all, we set $|z_1 - z_0| = \rho$. [As always, $\min\,(\alpha, \alpha, \beta) = \min\,(\alpha, \beta)$ and $\min\,(\alpha, \alpha) = \alpha$, and similarly with min replaced by max.]

rule for differentiating a quotient (see Rule 3, p. 108) that

$$f(z) = \frac{a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots}{b_0 + b_1(z - z_0) + \cdots + b_n(z - z_0)^n + \cdots} \tag{18.18}$$

is an analytic function on K. Therefore, according to Theorem 16.7, $f(z)$ has a power series representation

$$f(z) = c_0 + c_1(z - z_0) + \cdots + c_n(z - z_0)^n + \cdots \tag{18.19}$$

on K. By *division of power series* we mean the process of finding the "quotient series" (18.19) from a knowledge of the "dividend series" (18.16) and the "divisor series" (18.17). For further justification of this terminology, see Problem 18.8.

To find (18.19), we first combine (18.18) and (18.19), obtaining

$$\begin{aligned}[c_0 + c_1(z - z_0) + \cdots + c_n(z - z_0)^n + \cdots] \\ \times [b_0 + b_1(z - z_0) + \cdots + b_n(z - z_0)^n + \cdots] \\ = a_0 + a_1(z - z_0) + \cdots + a_n(z - z_0)^n + \cdots.\end{aligned} \tag{18.20}$$

Since both series in the left-hand side are absolutely convergent on K, they can be multiplied term by term (see Theorem 3.7). As a result, (18.20) becomes

$$\begin{aligned}c_0b_0 + (c_0b_1 + c_1b_0)(z - z_0) + (c_0b_2 + c_1b_1 + c_2b_0)(z - z_0)^2 + \cdots \\ + (c_0b_n + c_1b_{n-1} + \cdots + c_nb_0)(z - z_0)^n + \cdots \\ = a_0 + a_1(z - z_0) + a_2(z - z_0)^2 + \cdots + a_n(z - z_0)^n + \cdots,\end{aligned}$$

where both sides are convergent power series on K. It follows from Theorem 16.5 that identical powers of $z - z_0$ have identical coefficients. Thus we obtain an infinite system of linear equations

$$\begin{aligned}c_0b_0 &= a_0, \\ c_0b_1 + c_1b_0 &= a_1, \\ c_0b_2 + c_1b_1 + c_2b_0 &= a_2, \\ &\cdots\cdots\cdots \\ c_0b_n + c_1b_{n-1} + \cdots + c_nb_0 &= a_n, \\ &\cdots\cdots\cdots\end{aligned} \tag{18.21}$$

which determine the unknown coefficients c_n $(n = 0, 1, 2, \ldots)$.[5]

The problem of solving the system (18.21) is greatly facilitated by the fact that the first $n + 1$ equations $(n = 0, 1, 2, \ldots)$ involve only the first $n + 1$ unknowns, giving the system (18.21) its characteristic "triangular shape." Solving the first equation of the system, we obtain

$$c_0 = \frac{a_0}{b_0}$$

[5] The reader will recognize this technique as an instance of the method of undetermined coefficients, mentioned on p. 352.

($b_0 \neq 0$, by hypothesis). Substituting this value of c_0 into the second equation of the system, we obtain

$$c_1 = \frac{a_1 b_0 - a_0 b_1}{b_0^2},$$

and so on. Thus, having found $c_0, c_1, \ldots, c_{n-1}$, we substitute these values into the $(n+1)$th equation of the system (18.21), obtaining

$$c_n = \frac{a_n - c_0 b_n - c_1 b_{n-1} - \cdots - c_{n-1} b_1}{b_0}.$$

These successive substitutions can be avoided by using Cramer's rule, familiar from the theory of systems of linear equations.[6] The determinant of the system (18.21) is

$$\Delta = \begin{vmatrix} b_0 & 0 & 0 & 0 \\ b_1 & b_0 & 0 & 0 \\ b_2 & b_1 & b_0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ b_n & b_{n-1} & b_{n-2} & b_0 \end{vmatrix} = b_0^{n+1} \neq 0,$$

and hence

$$c_n = \frac{1}{b_0^{n+1}} \begin{vmatrix} b_0 & 0 & 0 & a_0 \\ b_1 & b_0 & 0 & a_1 \\ b_2 & b_1 & b_0 & a_2 \\ \cdot & \cdot & \cdot & \cdot \\ b_n & b_{n-1} & b_{n-2} & a_n \end{vmatrix} \quad (n = 0, 1, 2, \ldots). \quad (18.22)$$

Example 1. The function

$$f(z) = \frac{z}{e^z - 1} \quad (18.23)$$

is analytic everywhere except at those points where $e^z - 1$ vanishes and z does not, i.e., except at the points $\pm 2\pi i$, $\pm 4\pi i, \ldots$. Substituting the power series

$$e^z - 1 = \frac{z}{1!} + \frac{z^2}{2!} + \cdots + \frac{z^n}{n!} + \cdots$$

into (18.23), and dividing both numerator and denominator by z, we obtain

$$f(z) = \frac{1}{1 + \frac{z}{2!} + \cdots + \frac{z^n}{(n+1)!} + \cdots}$$

[6] See e.g., G. E. Shilov, *op. cit.*, Chap. 1, Sec. 7.

(note that this expression is meaningful for $z = 0$). The series in the denominator converges for all z and does not vanish for $z = 0$, but otherwise has the same zeros as $e^z - 1$. The two zeros closest to the origin are at $\pm 2\pi i$, and therefore $f(z)$ has a power series representation (18.19) on the disk $K: |z| < 2\pi$. To find this representation, we use our technique of division of power series. In the present case, the coefficients a_n, b_n appearing in the system (18.21) are just

$$a_0 = 1, \quad a_n = 0 \qquad (n = 1, 2, \ldots),$$
$$b_n = \frac{1}{(n+1)!} \qquad (n = 0, 1, 2, \ldots).$$

Therefore the first of the equations (18.21) gives

$$c_0 = 1,$$

and the rest reduce to the *recurrence relation*

$$c_0 \frac{1}{(n+1)!} + c_1 \frac{1}{n!} + \cdots + c_{n-1} \frac{1}{2!} + c_n = 0 \qquad (n = 1, 2, \ldots), \tag{18.24}$$

relating c_n to the values $c_0, c_1, \ldots, c_{n-1}$. Moreover, it follows from (18.22) that $c_0 = 1$ and[7]

$$c_n = \begin{vmatrix} 1 & 0 & 0 & 1 \\ \frac{1}{2!} & 1 & 0 & 0 \\ \frac{1}{3!} & \frac{1}{2!} & 1 & 0 \\ \cdot & \cdot & \cdot & 0 \\ \frac{1}{(n+1)!} & \frac{1}{n!} & \frac{1}{(n-1)!} & 0 \end{vmatrix}$$

$$= (-1)^n \begin{vmatrix} \frac{1}{2!} & 1 & 0 & 0 \\ \frac{1}{3!} & \frac{1}{2!} & 1 & 0 \\ \frac{1}{4!} & \frac{1}{3!} & \frac{1}{2!} & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \frac{1}{(n+1)!} & \frac{1}{n!} & \frac{1}{(n-1)!} & \frac{1}{2!} \end{vmatrix} \qquad (n = 1, 2, \ldots).$$

[7] To justify the transition from the first determinant to the second in the formula for c_n, see e.g., G. E. Shilov, *op. cit.*, Theorem 1, p. 11.

The numbers $c_n n!$ are called the *Bernoulli numbers* and are denoted by B_n. Thus $B_0 = c_0 0! = 1$ and

$$B_n = c_n n! = (-1)^n n! \begin{vmatrix} \frac{1}{2!} & 1 & 0 & 0 \\ \frac{1}{3!} & \frac{1}{2!} & 1 & 0 \\ \frac{1}{4!} & \frac{1}{3!} & \frac{1}{2!} & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \frac{1}{(n+1)!} & \frac{1}{n!} & \frac{1}{(n-1)!} & \frac{1}{2!} \end{vmatrix} \qquad (n = 1, 2, \ldots).$$

However, the easiest way to calculate the Bernoulli numbers is to use the recurrence relation (18.24), which now takes the form

$$B_0 \frac{1}{0!(n+1)!} + B_1 \frac{1}{1!n!} + \cdots + B_n \frac{1}{n!1!} = 0 \qquad (n = 1, 2, \ldots) \tag{18.25}$$

(recall that $B_0 = 1$). Multiplying (18.25) by $(n + 1)!$ and noting that

$$\frac{(n+1)!}{k!(n+1-k)!} = \binom{n+1}{k},$$

we obtain

$$B_0\binom{n+1}{0} + B_1\binom{n+1}{1} + \cdots + B_n\binom{n+1}{n} = 0 \qquad (n = 1, 2, \ldots). \tag{18.26}$$

Equation (18.26) can be written symbolically as

$$(1 + B)^{n+1} - B^{n+1} = 0, \tag{18.27}$$

where after raising $1 + B$ to the $(n + 1)$th power, every B^k is changed to B_k $(k = 1, 2, \ldots, n + 1)$. Using (18.27) and the fact that $B_0 = 1$, we deduce step by step that

$$\begin{aligned} B_0 + 2B_1 &= 0, & B_1 &= -\tfrac{1}{2}B_0 = -\tfrac{1}{2}, \\ B_0 + 3B_1 + 3B_2 &= 0, & B_2 &= -\tfrac{1}{3}B_0 - B_1 = \tfrac{1}{6}, \\ B_0 + 4B_1 + 6B_2 + 4B_3 &= 0, & B_3 &= -\tfrac{1}{4}B_0 - B_1 - \tfrac{3}{2}B_2 = 0, \\ B_0 + 5B_1 + 10B_2 + 10B_3 + 5B_4 &= 0, & & \\ & & B_4 &= -\tfrac{1}{5}B_0 - B_1 - 2B_2 - 2B_3 = -\tfrac{1}{30}, \\ B_0 + 6B_1 + 15B_2 + 20B_3 + 15B_4 + 6B_5 &= 0, & & \\ & & B_5 &= -\tfrac{1}{6}B_0 - B_1 - \tfrac{5}{2}B_2 - \tfrac{10}{3}B_3 - \tfrac{5}{2}B_4 = 0, \\ B_0 + 7B_1 + 21B_2 + 35B_3 + 35B_4 + 21B_5 + 7B_6 &= 0, & & \\ & & B_6 &= -\tfrac{1}{7}B_0 - B_1 - 3B_2 - 5B_3 - 5B_4 - 3B_5 = \tfrac{1}{42}, \\ & \cdots & & \end{aligned}$$

Collecting these results, we have

$$B_0 = 1, \quad B_1 = -\tfrac{1}{2}, \quad B_2 = \tfrac{1}{6}, \quad B_3 = 0, \quad B_4 = -\tfrac{1}{30}, \quad B_5 = 0, \quad B_6 = \tfrac{1}{42}, \ldots \tag{18.28}$$

Next we show that all the Bernoulli numbers with odd indices greater than 1 vanish. To see this, we write

$$f(z) = \frac{z}{e^z - 1} = c_0 + c_1 z + c_2 z^2 + c_3 z^3 + \cdots + c_n z^n + \cdots$$
$$= B_0 + \frac{B_1}{1!} z + \frac{B_2}{2!} z^2 + \frac{B_3}{3!} z^3 + \cdots + \frac{B_n}{n!} z^n + \cdots, \tag{18.29}$$

and then replace z by $-z$, obtaining

$$f(-z) = \frac{-z}{e^{-z} - 1} = \frac{-ze^z}{(e^{-z} - 1)e^z} = \frac{ze^z}{e^z - 1}$$
$$= B_0 - \frac{B_1}{1!} z + \frac{B_2}{2!} z^2 - \frac{B_3}{3!} z^3 + \cdots + \frac{B_n}{n!} (-1)^n z^n + \cdots. \tag{18.30}$$

Subtraction of (18.30) from (18.29) gives

$$f(z) - f(-z) = \frac{z}{e^z - 1} - \frac{ze^z}{e^z - 1} = -z$$
$$= 2\frac{B_1}{1!} z + 2\frac{B_3}{3!} z^3 + \cdots + 2\frac{B_{2m+1}}{(2m+1)!} z^{2m+1} + \cdots,$$

and hence, by the uniqueness of power series expansions,

$$2B_1 = -1, \quad B_3 = B_5 = \cdots = B_{2m+1} = \cdots = 0,$$

as asserted. Using this fact, we finally write the expansion (18.29) in the form

$$f(z) = \frac{z}{e^z - 1} = 1 - \frac{z}{2} + \sum_{m=1}^{\infty} \frac{B_{2m}}{(2m)!} z^{2m}, \tag{18.31}$$

where the series converges on the disk $K: |z| < 2\pi$. Moreover, the element $\{K, f(z)\}$ has singular points at $z = \pm 2\pi i$, since

$$\lim_{z \to \pm 2\pi i} \frac{z}{e^z - 1} = \infty,$$

and hence (18.31) has radius of convergence 2π. According to the Cauchy-Hadamard formula (16.2), this implies

$$\overline{\lim_{m \to \infty}} \sqrt[2m]{\frac{|B_{2m}|}{(2m)!}} = \frac{1}{2\pi},$$

since $B_{2m+1} = 0$ $(m = 1, 2, \ldots)$. Therefore, given any $\varepsilon > 0$, there exist infinitely many numbers B_{2m} satisfying the inequality

$$|B_{2m}| > \frac{(2m)!}{(2\pi + \varepsilon)^{2m}}.$$

If m is large, any such value of $|B_{2m}|$ greatly exceeds its index $2m$, despite the appearance of the first few Bernoulli numbers (18.28).

Example 2. Using the result of the preceding example, we can find the Taylor series expansions of $z \cot z$, $\tan z$ and $z \csc z$. First consider the function $z \cot z$. Since

$$\cot z = \frac{\cos z}{\sin z} = i\frac{e^{iz} + e^{-iz}}{e^{iz} - e^{-iz}} = i\frac{e^{2iz} + 1}{e^{2iz} - 1} = i + \frac{2i}{e^{2iz} - 1},$$

it follows that

$$z \cot z = iz + \frac{2iz}{e^{2iz} - 1}.$$

Replacing z by $2iz$ in (18.31), we find that

$$z \cot z = iz + 1 - \frac{2iz}{2} + \sum_{m=1}^{\infty} \frac{B_{2m}}{(2m)!}(2iz)^{2m} = 1 + \sum_{m=1}^{\infty} (-1)^m \frac{2^{2m}B_{2m}}{(2m)!} z^{2m}, \tag{18.32}$$

where the series converges for $|2iz| < 2\pi$, i.e., for $|z| < \pi$.

Next consider the function $\tan z$, and note that

$$\cot z - \tan z = 2 \cot 2z,$$

or

$$\tan z = \cot z - 2 \cot 2z.$$

Replacing z by $2z$ in (18.32), we find that

$$2z \cot 2z = 1 + \sum_{m=1}^{\infty} (-1)^m \frac{2^{4m}B_{2m}}{(2m)!} z^{2m}, \tag{18.33}$$

where the series converges for $|2z| < \pi$, i.e., for $|z| < \pi/2$. Subtraction of (18.33) from (18.32) gives

$$z \cot z - 2z \cot 2z = z \tan z = \sum_{m=1}^{\infty} (-1)^m \frac{2^{2m}(1 - 2^{2m})B_{2m}}{(2m)!} z^{2m},$$

or

$$\tan z = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{2^{2m}(2^{2m} - 1)B_{2m}}{(2m)!} z^{2m-1}, \tag{18.34}$$

where the series converges for $|z| < \pi/2$.

Finally, consider the function $z \csc z$, and note that

$$\cot z + \tan \frac{z}{2} = \frac{\cos z \cos \frac{z}{2} + \sin z \sin \frac{z}{2}}{\sin z \cos \frac{z}{2}} = \frac{\cos \frac{z}{2}}{\sin z \cos \frac{z}{2}} = \csc z.$$

Replacing z by $z/2$ in (18.34), we obtain a series which converges for

$|z/2| < \pi/2$, i.e., for $|z| < \pi$. Combining the result with the series (18.32), which also converges for $|z| < \pi$, we find that

$$\begin{aligned} z \csc z &= z \cot z + z \tan \frac{z}{2} \\ &= 1 + \sum_{m=1}^{\infty} (-1)^{m-1} \frac{(2^{2m} - 2)B_{2m}}{(2m)!} z^{2m} \end{aligned} \tag{18.35}$$

for $|z| < \pi$. It is easy to see that the series (18.32), (18.34) and (18.35) have radii of convergence π, $\pi/2$ and π, respectively, corresponding to the points nearest the origin at which the functions $z \cot z$, $\tan z$ and $z \csc z$ become infinite.

Example 3. The function

$$\sec z = \frac{1}{\cos z}$$

is analytic everywhere except at the zeros of $\cos z$, i.e., except at the points $z = \pm\pi/2, \pm 3\pi/2, \ldots$. The two zeros closest to the origin are at $z = \pm \pi/2$, and therefore $\sec z$ has a power series representation

$$\sec z = c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n + \cdots$$

which converges for $|z| < \pi/2$. Since $\sec z$ (like $\cos z$) is an even function, all the coefficients of odd powers of z vanish (cf. p. 350), and hence

$$\begin{aligned} \sec z &= \frac{1}{1 - \dfrac{z^2}{2!} + \dfrac{z^4}{4!} - \cdots + (-1)^n \dfrac{z^{2m}}{(2m)!} + \cdots} \\ &= c_0 + c_2 z^2 + c_4 z^4 + \cdots + c_{2m} z^{2m} + \cdots. \end{aligned} \tag{18.36}$$

Writing the coefficients of this series in the form

$$c_{2m} = (-1)^m \frac{E_{2m}}{(2m)!} \qquad (m = 0, 1, 2, \ldots),$$

we have

$$\sec z = \sum_{m=0}^{\infty} (-1)^m \frac{E_{2m}}{(2m)!} z^{2m}, \tag{18.37}$$

where the series has radius of convergence $\pi/2$.

To find the numbers E_{2m}, known as the *Euler numbers*, we observe that (18.36) implies

$$\begin{aligned} 1 = \left(1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \frac{z^8}{8!} - \cdots\right) \\ \times \left(E_0 - \frac{E_2 z^2}{2!} + \frac{E_4 z^4}{4!} - \frac{E_6 z^6}{6!} + \frac{E_8 z^8}{8!} - \cdots\right) \end{aligned}$$

or

$$1 = E_0 - \left(\frac{E_0}{2!} + \frac{E_2}{2!}\right)z^2 + \left(\frac{E_0}{4!} + \frac{E_2}{2!2!} + \frac{E_4}{4!}\right)z^4$$
$$- \left(\frac{E_0}{6!} + \frac{E_2}{4!2!} + \frac{E_4}{2!4!} + \frac{E_6}{6!}\right)z^6 + \left(\frac{E_0}{8!} + \frac{E_2}{6!2!} + \frac{E_4}{4!4!} + \frac{E_6}{2!6!} + \frac{E_8}{8!}\right)z^8 + \cdots$$

Therefore, by the uniqueness of power series expansions,

$$\begin{aligned}
E_0 &= 1,\\
\frac{E_0}{2!} + \frac{E_2}{2!} &= 0,\\
\frac{E_0}{4!} + \frac{E_2}{2!2!} + \frac{E_4}{4!} &= 0,\\
\frac{E_0}{6!} + \frac{E_2}{4!2!} + \frac{E_4}{2!4!} + \frac{E_6}{6!} &= 0,\\
\frac{E_0}{8!} + \frac{E_2}{6!2!} + \frac{E_4}{4!4!} + \frac{E_6}{2!6!} + \frac{E_8}{8!} &= 0,\\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot &
\end{aligned}$$

Solving these equations step by step, we find that

$$E_0 = 1, \quad E_2 = -1, \quad E_4 = 5, \quad E_6 = -61, \quad E_8 = 1385, \ldots$$

Clearly, the coefficients E_{2m} satisfy the recurrence relation

$$\frac{E_0}{(2m)!} + \frac{E_2}{(2m-2)!2!} + \frac{E_4}{(2m-4)!4!} + \cdots + \frac{E_{2m}}{(2m)!} = 0 \qquad (m = 1, 2, \ldots),$$

which becomes

$$E_0 + \binom{2m}{2}E_2 + \binom{2m}{4}E_4 + \cdots + \binom{2m}{2m-2}E_{2m-2} + E_{2m} = 0 \quad (m = 1, 2, \ldots), \tag{18.38}$$

when multiplied by $(2m)!$. It follows from (18.38) that E_{2m} is an integer if the numbers $E_0, E_2, \ldots, E_{2m-2}$ are integers, and hence that all the Euler numbers are integers, since $E_0 = 1$.

PROBLEMS

18.1. The function

$$\zeta(z) = \sum_{k=1}^{\infty} \frac{1}{k^z} = \sum_{k=1}^{\infty} e^{-z \ln k} \tag{18.39}$$

is known as the *Riemann zeta function*. Prove that the series (18.39) is uniformly convergent on every compact subset of the half-plane $\operatorname{Re} z > 1$, so that $\zeta(z)$ is analytic for $\operatorname{Re} z > 1$. Find the Taylor series expansion of $\zeta(z)$ at the point $z = 2$. What is the radius of convergence of this Taylor series?

18.2. Find the Taylor series expansion of

$$f(z) = \sum_{k=1}^{\infty} a_k \frac{z^k}{1 - z^k}$$

at the point $z = 0$. What is the radius of convergence of this Taylor series?

Hint. This result generalizes Example 1, p. 430, which corresponds to the case where all the $a_k = 1$.

18.3. By the *Euler phi function* $\phi(n)$, defined for every positive integer n, we mean the number of positive integers relatively prime to n (1 is considered to be relatively prime to itself).[8] Find the Taylor series expansion of

$$\sum_{k=1}^{\infty} \phi(k) \frac{z^k}{1 - z^k}$$

at the point $z = 0$. What is the radius of convergence of this Taylor series?

Hint. Use the result of the preceding problem and the fact that

$$\sum_{d \mid n} \phi(d) = n, \tag{18.40}$$

where the sum is over all positive integers d which divide n (including 1 and n).[9]

18.4. Why is the preliminary transformation

$$\exp \frac{1}{1 - z} = e \exp \frac{z}{1 - z}$$

made in Example 2, p. 435?

18.5. Find the terms up to degree 5 in z in the power series expansions of the function

a) $e^{z \sin z}$; b) $(1 + z)^z = e^{z \ln (1 + z)}$; c) e^{e^z}; d) $e^z \ln (1 + z)$.

18.6. Find the recurrence relation between the coefficients of the power series expansion of the function $\ln (1 + e^z)$, and prove that all the coefficients of odd powers of z greater than the first are equal to zero.

Hint. Note that

$$\ln (1 + e^z) - \ln (1 + e^{-z}) = z.$$

18.7. Given a power series

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \qquad (a_0 \neq 0),$$

---

[8] Thus $\phi(1) = 1$, $\phi(2) = 1$, $\phi(3) = 2$, $\phi(4) = 2$, $\phi(5) = 4, \ldots$

[9] For a proof of (18.40), see e.g., G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, third edition, Oxford University Press, London (1954), Chap. 5, Theorem 63.

with radius of convergence R, let z_0 be the zero of $f(z)$ which lies closest to the origin. Prove that

$$|z_0| \geqslant \frac{\rho|a_0|}{M(\rho) + |a_0|}$$

for any $\rho < R$, where

$$M(\rho) = \max_{|z| = \rho} |f(z)|.$$

Hint. If $|f(z) - a_0| < |a_0|$, then $f(z)$ cannot vanish. Use Cauchy's inequalities (16.43) to estimate $|f(z) - a_0|$.

18.8. Prove that formal "long division" of the series (18.16) by the series (18.17) gives a series (18.19) whose coefficients $c_0, c_1, c_2, \ldots, c_n, \ldots$ satisfy the system (18.21).

18.9. Find the Taylor series expansion at the origin, and the appropriate radius of convergence, of the function

a) $\ln \dfrac{\sin z}{z}$; b) $\ln \cos z$.

18.10. Prove that the coefficients c_n in the expansion

$$\frac{1}{1 - z - z^2} = \sum_{n=0}^{\infty} c_n z^n \tag{18.41}$$

satisfy the recurrence relation

$$c_0 = c_1 = 1, \quad c_n = c_{n-1} + c_{n-2} \qquad (n \geqslant 2).$$

What is the radius of convergence of the series (18.41)?

Comment. The numbers c_n are called the *Fibonacci numbers*.

18.11. In the expansion

$$\frac{A + Bz + Cz^2}{\alpha + \beta z + \gamma z^2 + \delta z^3} = \sum_{n=0}^{\infty} c_n z^n \qquad (\alpha \neq 0),$$

find c_0, c_1 and c_2, and also a recurrence relation between c_n, c_{n-1}, c_{n-2}, c_{n-3} $(n \geqslant 3)$.

BIBLIOGRAPHY

Ahlfors, L. V., *Complex Analysis. An Introduction to the Theory of Analytic Functions of One Complex Variable*, McGraw-Hill Book Co., New York (1953).

Bieberbach, L., *Conformal Mapping* (translated by F. Steinhardt), Chelsea Publishing Co., New York (1953).

Burkhardt, H., *Theory of Functions of a Complex Variable* (translated by S. E. Rasor), D. C. Heath and Co., Boston (1913).

Carathéodory, C., *Theory of Functions of a Complex Variable* (translated by F. Steinhardt), in two volumes, Chelsea Publishing Co., New York (1954).

Cartan, H., *Elementary Theory of Analytic Functions of One or Several Complex Variables*, Addison-Wesley Publishing Co., Inc., Reading, Mass. (1963).

Churchill, R. V., *Complex Variables and Applications*, second edition, McGraw-Hill Book Co., New York (1960).

Copson, E. T., *An Introduction to the Theory of Functions of a Complex Variable*, Oxford University Press, London (1935).

Dienes, P., *The Taylor Series. An Introduction to the Theory of Functions of a Complex Variable*, Dover Publications, Inc., New York (1957).

Estermann, T., *Complex Numbers and Functions*, University of London, The Athlone Press, London (1962).

Forsyth, A. R., *Theory of Functions of a Complex Variable*, Cambridge University Press, London (1918).

Franklin, P., *Functions of Complex Variables*, Prentice-Hall, Inc., Englewood Cliffs, N. J. (1958).

Fuchs, B. A. and V. I. Levin, *Functions of a Complex Variable and Some of Their Applications* (translated by J. Berry), Addison-Wesley Publishing Co., Inc., Reading, Mass. (1961). (This and the next book form a two-volume set.)

Fuchs, B. A. and B. V. Shabat, *Functions of a Complex Variable and Some of Their Applications* (translated by J. Berry and revised by J. W. Reed), Addison-Wesley Publishing Co., Inc., Reading, Mass. (1964).

Goursat, E., *Functions of a Complex Variable* (translated by E. R. Hedrick and O. Dunkel), Dover Publications, Inc., New York (1959).

Harkness, J. and F. Morley, *Introduction to the Theory of Analytic Functions*, G. E. Stechert and Co., New York (1924).

Hille, E., *Analytic Function Theory*, in two volumes, Ginn and Co., Boston (1959, 1962).

Knopp, K., *Elements of the Theory of Functions* (translated by F. Bagemihl), Dover Publications, Inc., New York (1952).

Knopp, K., *Theory of Functions* (translated by F. Bagemihl), in two volumes, Dover Publications, Inc., New York (1945, 1947).

LePage, W. R., *Complex Variables and the Laplace Transform for Engineers*, McGraw-Hill Book Co., New York (1961).

Littlewood, J. E., *Lectures on the Theory of Functions*, Oxford University Press, London (1944).

MacRobert, T. M., *Functions of a Complex Variable*, fourth edition, The Macmillan Co., New York (1958).

McLachlan, N. W., *Complex Variable Theory and Transform Calculus with Technical Applications*, second edition, Cambridge University Press, London (1955).

Miller, K. S., *Advanced Complex Calculus*, Harper and Row, Publishers, New York (1960).

Nehari, Z., *Conformal Mapping*, McGraw-Hill Book Co., New York (1952).

Nehari, Z., *Introduction to Complex Analysis*, Allyn and Bacon, Inc., Boston (1962).

Osgood, W. F., *Functions of a Complex Variable*, bound in one volume with *Functions of Real Variables* by the same author, Chelsea Publishing Co., New York (1958).

Pennisi, L. L., *Elements of Complex Variables* (with the collaboration of L. I. Gordon and S. Lasher), Holt, Rinehart and Winston, Inc., New York (1963).

Phillips, E. G., *Functions of a Complex Variable with Applications*, Oliver and Boyd, London (1961).

Pierpont, J., *Functions of a Complex Variable*, Dover Publications, Inc., New York (1959).

Saks, S. and A. Zygmund, *Analytic Functions* (translated by E. J. Scott), Nakładem Polskiego Towarzystwa Matematycznego, Warsaw (1952).

Samsone, G. and J. Gerretsen, *Lectures on the Theory of Functions of a Complex Variable. Volume I, Holomorphic Functions*, P. Noordhoff, Groningen (1960).

Schwerdtfeger, H., *Geometry of Complex Numbers*, University of Toronto Press, Toronto (1962).

Thron, W. J., *Introduction to the Theory of Functions of a Complex Variable*, John Wiley and Sons, Inc., New York (1953).

Titchmarsh, E. C., *The Theory of Functions*, second edition, Oxford University Press, London (1939).

Watson, G. N., *Complex Integration and Cauchy's Theorem*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 15, Cambridge University Press, London (1914), reprinted by Hafner Publishing Co., New York (1960).

Whittaker, E. T. and G. N. Watson, *A Course of Modern Analysis*, fourth edition, Cambridge University Press, London (1963).

Whyburn, G. T., *Topological Analysis*, Princeton University Press, Princeton, N. J. (1958).

INDEX

A

Abel's limit theorem, 396
Abel transformation, 38, 397
Aleksandrov, P. S., 70, 219
Analytic functions:
  of a complex variable, 9 ff., 112
    A-points of, 371
    at infinity, 125
    boundary values of, 370
      radial, 371
    circle average of, 295
    circular elements of, 382
    defined by integrals, 418–422
    families of, 411 ff.
      (conditionally) compact, 411
      equicontinuous, 427
    representation of, by infinite products, 338–339, 374–375
    zeros of, 372
  of a real variable, 3–7
Angle:
  interior of, 133
  of parallelism, 195
  with vertex at infinity, 90, 92
Anharmonic ratio (*see* Cross ratio)
Annulus, 285
Anti-derivative, 283
A-point, 371
  at infinity, 372
  multiple, 371
  order of, 371
  simple, 371
Arc, 60
  Jordan, 60
  simple, 60
Arc tangent, principal value of, 14
Archimedes' axiom, 192
Axiom(s):
  of congruence, 186
  of continuity, 192
  of incidence, 184
  of linear congruence, 188
  of order, 184
  of parallels, 192
  of planar congruence, 188–189

B

Bagemihl, F., 29
Bernoulli, John, 225
Bernoulli numbers, 440
  asymptotic behavior of, 441
  recurrence relation for, 440
Bernstein, S. N., 10
Binomial coefficient, 400
Binomial series, 365, 400, 408
Binomial theorem, 365
Birkhoff, G., 15, 16, 42, 168, 184, 186, 223
Bolzano-Weierstrass principle:
  for sequences, 28
  for sets, 44
Borsuk, K., 197
Boundary:
  of an arbitrary set, 76
  of an open set, 65
Boundary point:
  of an arbitrary set, 76
  of an open set, 65
Boundary uniqueness theorem, 371
Bouquet, J. C., 112
Branch point, 217 ff.
  algebraic, 217
  of finite order, 217
  of infinite order, 228

Branch point (*Cont.*)
at infinity, 217
logarithmic, 228
order of, 217
Briot, A. A., 112
Buck, R. C., 6, 37, 38, 40, 42, 114

C

Cantor's axiom of continuity, 192
Cauchy, A. L., 8, 112, 361
Cauchy convergence criterion, 31
Cauchy principal value, 290, 316, 342
Cauchy-Hadamard formula, 346
Cauchy-Hadamard inequality, 303, 363
Cauchy-Hadamard theorem, 344
Cauchy-Riemann equations, 110 ff.
generalization of, 117
in polar coordinates, 114
Cauchy's inequalities, 302
for Taylor coefficients, 358
Cauchy's integral, 293 ff.
Cauchy's integral formula, 293–295
for an unbounded domain, 319
Cauchy's integral theorem, 252–292
extension of, 270
modified versions of, 269
proof of, 261–268
for a system of contours, 279-281
used to evaluate definite integrals, 272–279
Cauchy's theorem on expansion of an analytic function in power series, 361
Circle of convergence, 346
Circular elements, 382 ff.
regular points of, 382
singular points of, 382
Closure:
of an arbitrary set, 75
of an open set, 66
Complex conjugate, 14
Compact families of analytic functions, 411–417
Complex number(s):
addition of, 14
geometric representation of, 16
arbitrary powers of, 228–231
argument of, 13
principal value of, 13
conjugate, 14
Complex number(s) (*Cont.*)
division of, 15
geometric representation of, 17
equality of, 12
geometric representation of, 12 ff.
imaginary, 12
imaginary part of, 12
improper, 78
modulus (absolute value) of, 13
motivation for introducing, 7
multiplication of, 14
geometric representation of, 17
*n*th power of, 17
*n*th root of, 18
principal value of, 19
polar form of, 140
proper, 78
purely imaginary, 12
rational powers of, 19
real part of, 12
sequences of (*see* Complex sequences)
series of (*see* Complex series)
subtraction of, 15
geometric representation of, 16
trigonometric form of, 13, 140
Complex plane, 13 ff.
cut along line segments, 240
extended, 85
compactness of, 85
conformal mapping of, 124
Complex sequence(s), 27 ff.
bounded, 28
convergence of moduli and arguments of, 33–35
convergence of real and imaginary parts of, 32
convergent, 29
converging to infinity, 78
coterminal, 77
in the extended plane, 84–85
limit of, 29
limit point of, 28
subsequences of, 27
uniformly bounded, 331
uniformly convergent, 330–334
Complex series, 35 ff.
absolutely convergent, 37
convergent, 36
divergent, 36
iterated, 41
partial sums of, 36

Complex series (*Cont.*)
sum of, 36
uniformly convergent (*see* Uniformly convergent series)
Complex variable, 8 ff.
Component:
of an arbitrary set, 75
of an open set, 64
Conformal mapping, 88, 120 ff.
of the extended plane, 124
of the first kind, 120
of the second kind, 120
Conjugate Fourier series, 357
Connected component (*see* Component)
Continuous curve (*see* Curve)
Continuum, 62, 85
Contour, 262
Contour integral, 249
Cosine, 146
modulus of, 148
real and imaginary parts of, 148
Courant, R., 144
Cramer's rule, 438
Cross ratio, 175, 190
invariance of, under Möbius transformations, 176
Curve:
closed, 60
end points of, 60
in the extended plane, 89
final point of, 60
initial point of, 60
inscribed, 246
Jordan, 60
length of, 246
multiple point of, 60
nonrectifiable, 246
normal to, 311
oriented, 60, 156
parameter of, 60
parametric equation of, 60
piecewise smooth, 247
points of, 60
polygonal, 61, 68
rectifiable, 246
simple closed, 60
smooth, 246
tangent to, 118
unbounded, 333

D

D'Alembert, J. le R., 111
De Moivre's theorem, 18
De Morgan formulas, 54
Density, 299
Dependent variable, 25
Derivative(s) of a complex function, 105
analyticity of, 301
geometric interpretation of, 118–122
Diagonal sequence, 413
Difference quotient, 105
Differential, 106
Disk, 66
Distance:
between a set and a point, 52
between two points, 16
between two sets, 52
Domain of definition, 24
Domain of univalence, 213
Domain(s), 64 ff.
approximation of, by an increasing sequence of compact sets, 373
closed, 66
finitely connected, 67, 85
infinitely connected, 67, 85
multiply connected, 67, 85
n-connected, 67, 85
properties inside, 411
simply connected, 67, 72, 85, 258

E

Efimov, N. V., 197
Elements (*see* Circular elements)
Entire functions, 128 ff.
transcendental, 135 ff.
Entire linear transformation, 122
ε-neighborhood, 27
Equator, 80
Equivalence relation, 186
Euler, L., 8, 111
Euler numbers, 443
recurrence relation for, 444
Euler phi function, 445
Euler's formula, 8, 140
Exponential, 135 ff.
addition theorem for, 135
to an arbitrary base, 231
Exterior:
of an arbitrary set, 76

Exterior (*Cont.*)
of a closed Jordan curve, 70
of an open set, 65
Exterior point:
of an arbitrary set, 76
of an open set, 65

F

Fatou's theorem, 404
Fibonacci numbers, 436
Finite arc, 333
Fixed point, 122, 171–173
Fractional linear transformation (*see* Möbius transformation)
Fresnel integrals, 272
Function(s), 24 ff.
analytic (*see* Analytic functions)
bounded, 50
of bounded variation, 254
complex, 25 ff.
of a complex variable, 25 ff.
composite, 47, 108, 433
continuous, 46 ff.
nowhere differentiable, 328
properties of, 47–52
uniformly, 51
continuous (*Cont.*)
wide-sense, 86
decreasing, 75
strictly, 75
differentiable, 105 ff.
complex, 105
real, 105
domain (of definition) of, 24
entire (*see* Entire functions)
even, 360
harmonic, 319
holomorphic, 112
increasing, 75
strictly, 75
infinitely differentiable, 5
infinite-valued, 224
inverse, 25, 109
limit of, 45
monotonic, 75
strictly, 75
multiple-valued, 24, 212–242
multivalent, 213
nondecreasing, 75
nonincreasing, 75
Function(s) (*Cont.*)
odd, 350
periodic, 140
range of, 24
regular, 112
rules for differentiation of, 107–110
schlicht, 213
simple, 213
single-valued, 24
univalent, 213
values of, 24
Fundamental theorem of algebra, 8, 128, 129, 377, 424
Fundamental theorem of integral calculus, 284

G

Gakhov, F. D., 316
Gamma function, 291, 428
Geometric series, 322, 325, 386, 395, 402, 404
Goffman, C., 310
Golubyev, V. V., 316
Green's theorem, 259
Group of transformations, 168

H

Half-strip, 72, 85, 154
Hardy, G. H., 403, 445
Heine-Borel property, 57
Heine-Borel theorem, 48
Hilbert, D., 184, 192, 379
Hilbert's axiom of completeness, 192
Hilbert's axioms, 184
Hilbert's lemniscate approximation theorem, 379, 380
Hille, E., 379
Homeomorphic image, 94
Homeomorphisms, 94–101
Homothetic transformation, 123
Hyperbolic cosine, 148
Hyperbolic functions, 148
inverse, 242
Hyperbolic sine, 148

I

Identity transformation (*see* Unit transformation)

Image:
of a point, 26, 27
of a set, 26
Imaginary axis, 13
Imaginary unit, 15
Inclination, 118, 199
Independent variable, 25
Infinite product(s), 334–340
absolutely convergent, 337, 343
convergent, 335, 340
divergent, 335
partial products of, 335
uniformly convergent, 338
value of, 335
Infinitely differentiable functions, 5
Initial meridian (*see* Prime meridian)
Inner measure, 370
Integral part, 304
Integrals of the Cauchy type, 299 ff.
boundary values of, 306–316
left-hand, 306
right-hand, 306
improper, 334
infinite differentiability of, 299
Integral(s) of complex functions, 248 ff.
Integral of complex functions (*Cont.*)
as a function of its upper limit in a multiply connected domain, 284-289
change of variable in, 304
examples of evaluation of, 252–253
indefinite, 283
path-independent, 282
properties of, 250–251
Interior:
of an arbitrary set, 76
of a closed Jordan curve, 70
of an open set, 65
Interior point:
of an arbitrary set, 76
of an open set, 65
Interior uniqueness theorem, 369
Intermediate value theorem, 75
Interval:
closed, 3
half-open, 332
open, 3
partition of, 245
Inverse image:
of a point, 26, 27
Inverse image (*Cont.*)
of a set, 26
Inversion in a circle, 179
Isolated point, 57
Isometric circle, 124
Isomorphism, 78

J

Jordan curve theorem, 70
Joukowski, N. Y., 197
Joukowski function, 197, 223

K

Key lemma, 259–261
Knopf, K., 29
Komm, H., 70

L

Lagrange, J. L., 112
Laplace transform, 422
Latitude, 80
Lemniscate of Bernoulli, 378
Lemniscates, 378–380
components of, 379
foci of, 378
radius of, 378
Limit inferior, 319
Limit point:
of a sequence, 28
of a set, 42
Limit superior, 303
Line at infinity, 84
Line integral, 248
Linear magnification ratio (*see* Magnification)
Liouville's theorem, 364
generalization of, 367
Littlewood, J. E., 403
Lobachevskian ($\mathscr{L}$-)angle, 188
exterior of, 188
interior of, 188
sides of, 188
size of, 190
vertex of, 188
$\mathscr{L}$-circle, 210
$\mathscr{L}$-congruence, 185
properties of, 186
$\mathscr{L}$-convexity, 188
$\mathscr{L}$-distance, 191

Lobachevskian (*Cont.*)
$\mathscr{L}$-geometry, 183–197
$\mathscr{L}$-length, 191
$\mathscr{L}$-lines, 184
$\mathscr{L}$-measure, 190
properties of, 190
$\mathscr{L}$-motion, 185
$\mathscr{L}$-parallels, 193
$\mathscr{L}$-plane, 184
points at infinity in, 192
$\mathscr{L}$-points, 184
$\mathscr{L}$-rays, 188
$\mathscr{L}$-triangle, 189
sides of, 189
$\mathscr{L}$-trigonometry, 196
Logarithm, 224 ff.
to an arbitrary base, 232
natural, 233
principal value of, 224
in terms of an indefinite integral, 287–289
Logarithmic series, 364, 395, 399, 408
Logarithmic spiral, 143
Longitude, 80
Lower limit, 319
Luzin, N. N., 371, 409
Luzin's series, 410–411

M

MacLane, S., 15, 16, 42, 168, 184, 186, 223
Magnification, 121
Majorant, 255
Mapping, 26 ff.
of a circle onto a circle, 176–177
continuous in both directions, 84, 94
"into," 26
one-to-one, 26
"onto," 26
topological, 94
$w = \text{Arc cos } z$, 234–237
$w = \cos z$, 150–157
$w = e^z$, 140–146
$w = \sqrt[n]{P_n(z)}$, 219–224
$w = \sqrt[n]{z}$, 214–219
$w = \frac{1}{2}(z + z^{-1})$, 197–202
$w = P_n(z)$, 130–132
$w = (z - a)^n$, 132–135
Mapping (*Cont.*)
$w = z + \ln z$, 237–239
$\zeta = 1/z$, 90–91
Markushevich, A. I., 316
Marquit, E., 197
Maximum (*see* Relative maximum)
Maximum modulus principle, 296, 298, 376
McShane, E. J., 144
Measure, 370
Meromorphic functions, 160–211
transcendental, 202
Method of undetermined coefficients, 352
Minimum modulus principle, 376
Möbius transformation(s), 122 ff.
associativity of multiplication of, 167
circle-preserving property of, 168–171
closure of multiplication of, 167–168
conformality of, 123, 125
determinant of, 165
fixed points of, 122, 171–172
group property of, 165
products of, 166
symmetry-preserving property of, 178
Modular surface, 149
of $\sin z$, 149
of $\tan z$, 205
Montel's compactness criterion, 415
Morera's theorem, 301
Multiple-valued functions, 24, 212–242
Multiplicity, 129
Muskhelishvili, N. I., 316

N

Neighborhood, 3, 27 ff.
half-size, 51, 53
on the Riemann sphere, 82
Nontangential path, 311
Normal to a curve, 311
North pole, 80
as image of infinity, 83

O

One-to-one correspondence, 26
Open set:
boundary of, 65
boundary points of, 65
closure of, 66

Open set (*Cont.*)
(connected) components of, 64
exterior of, 65
exterior points of, 65
interior of, 65
interior points of, 65
Orthogonal system, 134
Orthogonality relations, 354
Outer measure, 370
Oval of Cassini, 378

P

Parseval's theorem, 355
Partition, 245
norm of, 245
Path of integration, 249
Path-independent integral, 282
Period, 140
fundamental (primitive), 140, 146
Plemelj, J., 316
Plemelj formulas, 316
Point at infinity, 84
Pole of a rational function, 161
at infinity, 161
order of, 161
Polynomial(s), 128–135
coefficients of, 128
degree of, 128
zeros of, 129
Power series, 344 ff.
behavior of, on circle of convergence, 395–411
division of, 436–444
expansion of an analytic function in, 358–365
formal substitution of, into another power series, 433
relation of, to Fourier series, 353–358
Pre-image (*see* Inverse image)
Prime meridian, 80
Primitive, 283
Pringsheim's theorem, 389
Privalov, I. I., 316, 371
Projective plane, 84

R

Radius of convergence, 346
Radok, J. R. M., 316
Ratio test, 38
Rational function(s), 160–165
order of, 162
poles of, 161
zeros of, 161
Ray center, 123
Ray ratio, 123
Real axis, 12
Recurrence relation, 439
Reflection in a straight line or circle, 179
Region, 159
Regular points, 382
Relative maximum:
in the strict sense, 296
in the wide sense, 296
Riemann, B., 79
Riemann sphere, 79
hemispheres of, 80
sets of points on, 82
Riemann zeta function, 444
Root of an equation, 129
multiplicity of, 129
Root test, 38

S

Schwarz's lemma, 381
Series with complex terms (*see* Complex series)
Set inclusion, 23
Set(s), 23 ff.
arcwise connected, 62
bounded, 44
closed, 47
compact, 48, 57, 411
complement of, 23
conditionally compact, 411
connected, 59
convex, 74
countable, 55
countably infinite, 55
dense in itself, 324
derived, 59
diameter of, 58
difference of, 24
disconnected, 59
disjoint, 23
distance between, 52–54
empty, 23
equivalent, 54
finite, 54

Set(s) (*Cont.*)
of functions uniformly bounded inside a domain, 415
homeomorphic, 94
infinite, 54
intersection of, 23
measurable, 370
of measure zero, 310
member of, 23
nonempty, 23
open (*see* Open set)
pairwise disjoint, 24
uncountable, 55
union of, 23
Shilov, G. E., 77, 123, 438, 439
Silverman, R. A., 77, 168, 355
Sine, 146
modulus of, 148
real and imaginary parts of, 148
Single-valued branch, 115, 212–213
Singular points, 3, 5, 382
everywhere dense set of, 387
Smielew, W., 197
Smirnov, V. I., 168, 184
Sokhotski, Y. V., 316
Sokhotski formulas, 316
South pole, 80
Square(s), 28, 71, 95
center of, 28, 95
chained, 71
Stereographic projection, 79 ff.
circle-preserving property of, 82
conformality of, 88
Stieltjes integral, 255
Stirling's formula, 346
Strip, 72, 85, 90, 143
Subgroup, 184, 207
Subset, 23
proper, 23
Symmetry transformations, 174–181

T

Tangent to a curve, 118
inclination of, 118
left-hand, 126
right-hand, 126
Tangent vector, 119
Tauber's theorem, 402
Taylor coefficients, 358
Taylor series, 349 ff.
of a composite function, 432–436
Taylor series (*Cont.*)
of the sum of a series of analytic functions, 429–432
Titchmarsh, E. C., 328, 404
Tolstov, G. P., 355
Townsend, E. J., 184
Transformation, 26
of similitude, 123
topological, 94
Trigonometric functions, 146, 202–207
addition theorems for, 147
inverse, 242
reduction formulas for, 147
Trigonometric polynomial, 203
degree of, 203

U

Uniformly bounded sequences, 331
Uniformly convergent families, 332
Uniformly convergent sequences, 330–334
Uniformly convergent series, 321–329
term-by-term integration of, 325
indefinite, 341
Uniqueness property, 350
Uniqueness theorems, 350, 369–371
Unit transformation, 166
Univalence, domains of, 213
Upper limit, 303

V

Vitali's theorem, 417

W

Weierstrass, K., 112, 328
Weierstrass' M-test, 323
Weierstrass' theorem:
on uniformly convergent families of analytic functions, 333
on uniformly convergent sequences of analytic functions, 330
on uniformly convergent series of analytic functions, 326
Widder, D. V., 6, 11, 110, 113, 255, 259, 291, 346
Wide-sense continuity, 86
Wolfe, H. E., 197
Wright, E. M., 445

Z

Zero-points (*see* Zeros)
Zeros:
  of analytic functions, 372
  of polynomials, 87, 129
    order of, 129
Zeros (*Cont.*)
  of rational functions, 161
    at infinity, 161
    multiple, 161
    order of, 161
    simple, 161

CONTENTS OF VOLUME I

PART 1. BASIC CONCEPTS

Chap. 1. Introduction
Chap. 2. Complex Numbers
Chap. 3. Sets and Functions. Limits and Continuity
Chap. 4. Connectedness, Curves and Domains
Chap. 5. Infinity and Sterographic Projection
Chap. 6. Homeomorphisms

PART 2. DIFFERENTIATION. ELEMENTARY FUNCTIONS

Chap. 7. Differentiation and the Cauchy-Riemann Equations
Chap. 8. Geometric Interpretation of the Derivative. Conformal Mapping
Chap. 9. Elementary Entire Functions
Chap. 10. Elementary Meromorphic Functions
Chap. 11. Elementary Multiple-Valued Functions

PART 3. INTEGRATION. POWER SERIES

Chap. 12. Rectifiable Curves. Complex Integrals
Chap. 13. Cauchy's Integral Theorem
Chap. 14. Cauchy's Integral and Related Topics
Chap. 15. Uniform Convergence. Infinite Products
Chap. 16. Power Series: Rudiments
Chap. 17. Power Series: Ramifications
Chap. 18. Methods for Expanding Functions in Taylor Series

CONTENTS OF VOLUME II

PART 1. LAURENT SERIES. CALCULUS OF RESIDUES

Chap. 1. Laurent Series. Isolated Singular Points
Chap. 2. The Calculus of Residues and Its Applications
Chap. 3. Inverse and Implicit Functions
Chap. 4. Univalent Functions

PART 2. HARMONIC AND SUB-HARMONIC FUNCTIONS

Chap. 5. Basic Properties of Harmonic Functions
Chap. 6. Applications to Fluid Dynamics
Chap. 7. Subharmonic Functions
Chap. 8. The Poisson-Jensen Formula and Related Topics

PART 3. ENTIRE AND MEROMORPHIC FUNCTIONS

Chap. 9. Basic Properties of Entire Functions
Chap. 10. Infinite Product and Partial Fraction Expansions

SELECTED RUSSIAN PUBLICATIONS IN THE MATHEMATICAL SCIENCES

RICHARD A. SILVERMAN, *Editor*

Formerly of the *Courant Institute, New York University*

AN INTRODUCTION TO THE THEORY OF LINEAR SPACES
by G. E. SHILOV
Published 1961 — *310 pages*

FOURIER SERIES
by G. P. TOLSTOV
Published 1962 — *336 pages*

EXTRACTION OF SIGNALS FROM NOISE
by L. A. WAINSTEIN and V. D. ZUBAKOV
Published 1962 — *382 pages*

AN INTRODUCTION TO THE THEORY OF STATIONARY RANDOM FUNCTIONS
by A. M. YAGLOM
Published 1962 — *235 pages*

CALCULUS OF VARIATIONS
by I. M. GELFAND and S. V. FOMIN
Published 1963 — *232 pages*

METHODS OF QUANTUM FIELD THEORY IN STATISTICAL PHYSICS
by A. A. ABRIKOSOV, L. P. GORKOV and I. E. DZYALOSHINSKI
Published 1963 — *352 pages*

SPECIAL FUNCTIONS AND THEIR APPLICATIONS
by N. N. LEBEDEV
Published 1965 — *308 pages*

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE, Volume I
by A. I. MARKUSHEVICH
Published 1965 — *459 pages*

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE, Volume II
by A. I. MARKUSHEVICH
Published 1965 — *333 pages*

PROBLEMS OF MATHEMATICAL PHYSICS
by N. N. LEBEDEV, I. P. SKALSKAYA and Y. S. UFLYAND
Published 1965 — *429 pages*

INTEGRAL, MEASURE AND DERIVATIVE: A UNIFIED APPROACH
by G. E. SHILOV and B. L. GUREVICH
Published 1966 — *233 pages*

ORDINARY DIFFERENTIAL EQUATIONS
by I. G. PETROVSKI
Published 1966 — *232 pages*